

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 21:52:49 ; Search time 2196 Seconds

(without alignments)
10085.262 Million cell updates/sec

Title: US-10-024-955-6

Sequence: 1 GATCTATATCATACATACAT.....ATTTTATTTTATTTCTCC 761

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

```
GenEmbl: *
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.ov:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.ph:*
24: em.pl:*
25: em.ro:*
26: em.sts:*
27: em.un:*
28: em.vi:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.in:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pin:*
35: em.htg.rod:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB ID	Description
	Score	Match	Length		
1	761	100.0	761	3 S80655	S80655 allergen Der
2	761	100.0	761	6 AX028702	AX028702 Sequence
3	47	6.2	812	3 DP037044	DP037044 Dermatophag
4	47	6.2	812	6 AR098569	AR098569 Sequence
5	47	6.2	812	6 AX028697	AX028697 Sequence
6	24	3.2	24	6 AX028708	AX028708 Sequence
7	24	3.2	183515	6 AC107223	AC107223 Homo sapi
8	23	3.0	2201	8 CG236287	AJ236287 Carphalea
9	23	3.0	122517	2 AC091350	AC091350 Rattus no
10	23	3.0	164131	2 AC129204	AC129204 Mus muscu
11	23	3.0	196029	2 AC129100	AC129100 Rattus no
12	23	3.0	208590	2 AC122636	AC122636 Rattus no
13	22	2.9	11473	6 AX346256	AX346256 Sequence
14	22	2.9	11473	6 AX348416	AX348416 Sequence
15	22	2.9	46387	2 U82212	U82212 Homo sapien
16	22	2.9	66591	9 AL589910	AL589910 Human DNA
17	22	2.9	97451	2 AC068603	AC068603 Homo sapi
18	22	2.9	116552	2 AC004022	AC004022 Homo sapi
19	22	2.9	122972	2 AL691434	AL691434 Homo sapi
20	22	2.9	14891	9 AL162274	AL162274 Human DNA
21	22	2.9	151136	2 HSA189K21	AL121893 Human DNA
22	22	2.9	154716	8 AC021043	AC021043 Atribidops
23	22	2.9	154850	9 AC055711	AC055711 Homo sapi
24	22	2.9	158882	2 AL732407	AL732407 Homo sapi
25	22	2.9	160295	10 AC122884	AC122884 Mus muscu
26	22	2.9	163838	2 AC069241	AC069241 Homo sapi
27	22	2.9	165120	2 AC069430	AC069430 Homo sapi
28	22	2.9	167837	2 AL732395	AL732395 Homo sapi
29	22	2.9	180915	8 GTAJ10592	AJ010592 Guillardia
30	22	2.9	182679	9 AL365207	AL365207 Human DNA
31	22	2.9	183594	2 AC122485	AC122485 Mus muscu
32	22	2.9	249289	2 AC103310	AC103310 Rattus no
33	22	2.8	28	6 AX028709	AX028709 Sequence
34	21	2.8	493	3 AF289137	AF289137 Aphelinus
35	21	2.8	493	3 AF289138	AF289138 Aphelinus
36	21	2.8	509	8 AB032753	AB032753 Solanum m
37	21	2.8	2171	3 GRI302036	AJ302036 Gryllus b
38	21	2.8	6005	1 AF078038S5	AF078042 Caldicell
39	21	2.8	6416	6 E35100	E35100 Truncated c
40	21	2.8	6725	6 AX281444	AX281444 Sequence
41	21	2.8	6725	6 AX346111	AX346111 Sequence
42	21	2.8	9760	6 AX251852	AX251852 Sequence
43	21	2.8	9760	6 AX344238	AX344238 Sequence
44	21	2.8	9760	6 AX348629	AX348629 Sequence
45	21	2.8	15360	3 AF346733	AF346733 Tetrahyme

ALIGNMENTS

RESULT 1
S80655
LOCUS
DEFINITION allergen Der f 7 [Dermatophagoides pteronyssinus-house dust mites,
MRNA, 761 nt].
ACCESSION S80655
VERSION S80655
KEYWORDS
SOURCE MRNA, 761 nt].
ORGANISM 761 bp
Dermatophagoides pteronyssinus.
Dermatophagoides pteronyssinus.
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcoptiformes; Ascligmata; Analgoidea; Pyroglyphidae;
Dermatophagoides.
REFERENCE
1 (Bases 1 to 761)
Shen,H.D., Chua,K.Y., Lin,W.L., Hsieh,K.H. and Thomas,W.R.

TITLE Molecular cloning and immunological characterization of the house dust mite allergen Der f 7
JOURNAL Clin. Exp. Allergy 25 (10), 1000-1006 (1995)
MEDLINE 96120794
PubMed 8556554
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI g1bbsq 174942] from the original journal article.
This sequence comes from Fig. 1.
Location/Qualifiers
FEATURES
source
1..761
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
gene
1..761
/gene="allergen Der f 7"
43..684
/gene="allergen Der f 7"
/note="this sequence comes from Fig. 1"
CDS
/codon_start=1
/product="allergen Der f 7"
/protein_id="AAB3597.1"
/db_xref="GI:1311689"
/translation="MKFLLIAVAFAVVSADPIHYDKITEERINKAIDATAIAIPOSE
TIDPKVDPDHADKPERHVGIVDFKGLAMRNTEARGLKOMKROGDANVKGEGIVKAH
LLIGVHDIIVSEMYDLATKLDLHPTTHVYSIDQPVVALSIEISDEGNITHTSEVR
QFANVNHIGLSIIDPIFGVLSVDLTAFIDTVKREMTKVLAFAFKRELEKN"
BASE COUNT 245 a 134 c 137 g 245 t
ORIGIN
Query Match 100.0%; Score 761; DB 3; Length 761;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCTATATTCATTAACAATCCAAAAAACAATATCTTACAAAATGATGAATTTTGTG 60
DB 1 GATCTATATTCATTAACAATCCAAAAAACAATATCTTACAAAATGATGAATTTTGTG 60
QY 61 ATTGCTGCCGTGGCATTTTGTGCGCGTTTGGCTGATCAATTCATCATGATAAATCACC 120
DB 61 ATTGCTGCCGTGGCATTTTGTGCGCGTTTGGCTGATCAATTCATCATGATAAATCACC 120
QY 121 GAAGAATCAACAAGCTATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 121 GAAGAATCAACAAGCTATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 181 GATCCAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGGTATGTGAT 240
DB 181 GATCCAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGGTATGTGAT 240
QY 241 TTCAAAGTGAATTTACCATGCGAACAATTGAGGCTCGAGGATGTAACAATGAACG 300
DB 241 TTCAAAGTGAATTTACCATGCGAACAATTGAGGCTCGAGGATGTAACAATGAACG 300
QY 301 CAAGGTGATGCTAATGTCGAAGGTGAAGGATGTTGAAGCTCATTTGTTGATCGGT 360
DB 301 CAAGGTGATGCTAATGTCGAAGGTGAAGGATGTTGAAGCTCATTTGTTGATCGGT 360
QY 361 GTTCACGATGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 GTTCACGATGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 CCACACGATGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 421 CCACACGATGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 481 TCTGATGAAGGTGAACATTAACATGATGATGATGATGATGATGATGATGATGATGAT 540
DB 481 TCTGATGAAGGTGAACATTAACATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 541 AACCATATTTGGTGTCTTCAATCTTGATCAATTTTGGCTTTATCTGATGATGAT 600
DB 541 AACCATATTTGGTGTCTTCAATCTTGATCAATTTTGGCTTTATCTGATGATGAT 600
QY 601 ACCGCTATTTTCCAGAACCCGTAGCTAAGGAATGACCAAGATATTGGCACCAGCATTT 660

DB 601 ACCGCTATTTTCCAGAACCCGTAGCTAAGGAATGACCAAGATATTGGCACCAGCATTT 660
QY 661 AAACGTGAATGGAAAAAATTAACCAATAGACATCATTTTCCAAAGTACATCTCTA 720
DB 661 AAACGTGAATGGAAAAAATTAACCAATAGACATCATTTTCCAAAGTACATCTCTA 720
QY 721 TTTCACGATCAATTAATAATTTTATTTTATTTTCTCC 761
DB 721 TTTCACGATCAATTAATAATTTTATTTTATTTTCTCC 761
RESULT 2
AX028702 761 bp DNA linear PAT 16-SEP-2000
LOCUS
DEFINITION Sequence 6 from Patent EP1018550.
ACCESSION AX028702
VERSION AX028702.1 GI:10189815
KEYWORDS
SOURCE
ORGANISM
European house dust mite.
Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;
Dermatophagoides.
REFERENCE
1 (bases 1 to 761)
AUTHORS Thomas, W.R. and Chua, K.Y.
TITLE Allergenic protein and peptides from house dust mite and uses thereof
PATENT: EP 1018550-A 6 12-JUL-2000;
JOURNAL INST CHILD HEALTH RESEARCH (AU)
FEATURES
source
1..761
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
43..684
/note="unnamed protein product"
CDS
/codon_start=1
/protein_id="CAC09235.1"
/db_xref="GI:10189816"
/translation="MKFLLIAVAFAVVSADPIHYDKITEERINKAIDATAIAIPOSE
TIDPKVDPDHADKPERHVGIVDFKGLAMRNTEARGLKOMKROGDANVKGEGIVKAH
LLIGVHDIIVSEMYDLATKLDLHPTTHVYSIDQPVVALSIEISDEGNITHTSEVR
QFANVNHIGLSIIDPIFGVLSVDLTAFIDTVKREMTKVLAFAFKRELEKN"
BASE COUNT 245 a 134 c 137 g 245 t
ORIGIN
Query Match 100.0%; Score 761; DB 6; Length 761;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCTATATTCATTAACAATCCAAAAAACAATATCTTACAAAATGATGAATTTTGTG 60
DB 1 GATCTATATTCATTAACAATCCAAAAAACAATATCTTACAAAATGATGAATTTTGTG 60
QY 61 ATTGCTGCCGTGGCATTTTGTGCGCGTTTGGCTGATCAATTCATCATGATAAATCACC 120
DB 61 ATTGCTGCCGTGGCATTTTGTGCGCGTTTGGCTGATCAATTCATCATGATAAATCACC 120
QY 121 GAAGAATCAACAAGCTATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 121 GAAGAATCAACAAGCTATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 181 GATCCAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGGTATGTGAT 240
DB 181 GATCCAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGGTATGTGAT 240
QY 241 TTCAAAGTGAATTTACCATGCGAACAATTGAGGCTCGAGGATGTAACAATGAACG 300
DB 241 TTCAAAGTGAATTTACCATGCGAACAATTGAGGCTCGAGGATGTAACAATGAACG 300
QY 301 CAAGGTGATGCTAATGTCGAAGGTGAAGGATGTTGAAGCTCATTTGTTGATCGGT 360
DB 301 CAAGGTGATGCTAATGTCGAAGGTGAAGGATGTTGAAGCTCATTTGTTGATCGGT 360

mat_peptide 119..712
/product="Der p 7 allergen polypeptide"
misc_signal 518..526
/note="N-glycosylation signal"
polya_signal 765..770
BASE COUNT 270 a 132 c 143 g 267 t
ORIGIN

Query Match 6.2%; Score 47; DB 3; Length 812;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 128
|||||
Db 107 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 153
|||||

RESULT 4
LOCUS AR098569 812 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6077517.
ACCESSION AR098569
VERSION AR098569.1 GI:12808335
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas, W. Robert, and Chua, K.-Y.
TITLE House dust mite allergen, Der p VII, and uses thereof
JOURNAL Patent: US 6077517-A 1 20-JUN-2000;
FEATURES
LOCATION/Qualifiers
SOURCE 1..812
/organism="unknown"
BASE COUNT 270 a 133 c 142 g 267 t
ORIGIN

Query Match 6.2%; Score 47; DB 6; Length 812;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 128
|||||
Db 107 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 153
|||||

RESULT 5
LOCUS AX028697 812 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent EP1018550.
ACCESSION AX028697
VERSION AX028697.1 GI:10189810
KEYWORDS
SOURCE European house dust mite.
ORGANISM Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Pyroglyphidae;
Dermatophagoides.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas, W. R. and Chua, K. Y.
TITLE Allergenic protein and peptides from house dust mite and uses thereof
JOURNAL Patent: EP 1018550-A 1 12-JUL-2000;
FEATURES
LOCATION/Qualifiers
SOURCE 1..812
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
/clone="H06"
68..715
/codon_start=1
/product="Der p 7 allergen polypeptide"
/protein_id="AA80264.1"
/db_xref="GI:1045602"
/translation="MMKLILIAAFAVASDPIHYDKITPEINKKAVDEAVATIEKSE
TFDPKVPDHSKDFEIRHIGITDLKGEIDMRNIQVRGLKMRVGDANVKSDDGVKKAH
LVGVHDDVYSMEYDLAKIKGLDLPNTHVLSIDDFVVELSEVSEGNMTLTSFEVR
QFANVNHIGLSILDPIFAVLSDVLAIFODTVRAEMTKVLAFAFKKELRNNO"
68..118
/note="putative leader peptide"

mat_peptide 119..712
/product="Der p 7 allergen polypeptide"
misc_signal 518..526
/note="N-glycosylation signal"
polya_signal 765..770
BASE COUNT 270 a 132 c 143 g 267 t
ORIGIN

Query Match 6.2%; Score 47; DB 3; Length 812;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 128
|||||
Db 107 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 153
|||||

RESULT 4
LOCUS AR098569 812 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6077517.
ACCESSION AR098569
VERSION AR098569.1 GI:12808335
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas, W. Robert, and Chua, K.-Y.
TITLE House dust mite allergen, Der p VII, and uses thereof
JOURNAL Patent: US 6077517-A 1 20-JUN-2000;
FEATURES
LOCATION/Qualifiers
SOURCE 1..812
/organism="unknown"
BASE COUNT 270 a 133 c 142 g 267 t
ORIGIN

Query Match 6.2%; Score 47; DB 6; Length 812;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 128
|||||
Db 107 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 153
|||||

RESULT 5
LOCUS AX028697 812 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent EP1018550.
ACCESSION AX028697
VERSION AX028697.1 GI:10189810
KEYWORDS
SOURCE European house dust mite.
ORGANISM Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Pyroglyphidae;
Dermatophagoides.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas, W. R. and Chua, K. Y.
TITLE Allergenic protein and peptides from house dust mite and uses thereof
JOURNAL Patent: EP 1018550-A 1 12-JUL-2000;
FEATURES
LOCATION/Qualifiers
SOURCE 1..812
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
/clone="H06"
68..715
/codon_start=1
/product="Der p 7 allergen polypeptide"
/protein_id="AA80264.1"
/db_xref="GI:1045602"
/translation="MMKLILIAAFAVASDPIHYDKITPEINKKAVDEAVATIEKSE
TFDPKVPDHSKDFEIRHIGITDLKGEIDMRNIQVRGLKMRVGDANVKSDDGVKKAH
LVGVHDDVYSMEYDLAKIKGLDLPNTHVLSIDDFVVELSEVSEGNMTLTSFEVR
QFANVNHIGLSILDPIFAVLSDVLAIFODTVRAEMTKVLAFAFKKELRNNO"
68..118
/note="putative leader peptide"

mat_peptide 119..712
/product="Der p 7 allergen polypeptide"
misc_signal 518..526
/note="N-glycosylation signal"
polya_signal 765..770
BASE COUNT 270 a 132 c 143 g 267 t
ORIGIN

Query Match 6.2%; Score 47; DB 3; Length 812;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 128
|||||
Db 107 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 153
|||||

RESULT 4
LOCUS AR098569 812 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6077517.
ACCESSION AR098569
VERSION AR098569.1 GI:12808335
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas, W. Robert, and Chua, K.-Y.
TITLE House dust mite allergen, Der p VII, and uses thereof
JOURNAL Patent: US 6077517-A 1 20-JUN-2000;
FEATURES
LOCATION/Qualifiers
SOURCE 1..812
/organism="unknown"
BASE COUNT 270 a 133 c 142 g 267 t
ORIGIN

Query Match 6.2%; Score 47; DB 6; Length 812;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 128
|||||
Db 107 GCCGTTTCGGCTGATCCATTCATGATGATAAATCACCAGGAANAAT 153
|||||

RESULT 5
LOCUS AX028697 812 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent EP1018550.
ACCESSION AX028697
VERSION AX028697.1 GI:10189810
KEYWORDS
SOURCE European house dust mite.
ORGANISM Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Pyroglyphidae;
Dermatophagoides.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas, W. R. and Chua, K. Y.
TITLE Allergenic protein and peptides from house dust mite and uses thereof
JOURNAL Patent: EP 1018550-A 1 12-JUL-2000;
FEATURES
LOCATION/Qualifiers
SOURCE 1..812
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
/clone="H06"
68..715
/codon_start=1
/product="Der p 7 allergen polypeptide"
/protein_id="AA80264.1"
/db_xref="GI:1045602"
/translation="MMKLILIAAFAVASDPIHYDKITPEINKKAVDEAVATIEKSE
TFDPKVPDHSKDFEIRHIGITDLKGEIDMRNIQVRGLKMRVGDANVKSDDGVKKAH
LVGVHDDVYSMEYDLAKIKGLDLPNTHVLSIDDFVVELSEVSEGNMTLTSFEVR
QFANVNHIGLSILDPIFAVLSDVLAIFODTVRAEMTKVLAFAFKKELRNNO"
68..118
/note="putative leader peptide"

```

mat_peptide 270 a 133 c 142 g 267 t
BASE COUNT
ORIGIN
Query Match 6.2%; Score 47; DB 6; Length 812;
Best Local Similarity 100.0%; Pred. No. 8e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 GCCGTTGGCGTCAATTCACATGATGAATCAACGAGAAAT 128
Db 107 GCCGTTGGCGTCAATTCACATGATGAATCAACGAGAAAT 153

RESULT 6
AX028708 AX028708 24 bp DNA linear PAT 16-SEP-2000
LOCUS Sequence 12 from Patent EP1018550.
ACCESSION AX028708
VERSION AX028708.1 GI:10189821
KEYWORDS
SOURCE European house dust mite.
ORGANISM Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;
Dermatophagoides.
REFERENCE 1 (bases 1 to 24)
AUTHORS Thomas, W.R. and Chua, K.Y.
TITLE Allergenic protein and peptides from house dust mite and uses
JOURNAL Patent: EP 1018550-A 12-12-JUL-2000;
FEATURES INST CHILDR HEALTH RESEARCH (AU)
Source Location/Qualifiers
1.24
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
BASE COUNT 5 a 5 c 4 g 10 t
ORIGIN
Query Match 3.2%; Score 24; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 559 TCAATCTTGATCCCAATTTTGGC 582
Db 1 TCAATCTTGATCCCAATTTTGGC 24

RESULT 7
AC107223/c AC107223 183515 bp DNA linear PRI 20-APR-2002
LOCUS Homo sapiens BAC clone RP11-1289C17 from 4, complete sequence.
DEFINITION AC107223
ACCESSION AC107223
VERSION AC107223.5 GI:20136991
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183515)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
AUTHORS 2 (bases 1 to 183515)
REFERENCE Bielicki, L., Kruchowski, S., Cotton, M. and Doeber, A.
TITLE The sequence of Homo sapiens BAC clone RP11-1289C17
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 183515)

```

```

AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 183515)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 183515)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 183515)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 11, 2002 this sequence version replaced gi:18767625.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH1289C17
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPI1-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-5005; the clone sequenced
to the right is RP13-93318. Actual start of this clone is at base
position 1 of RP11-1289C17; actual end is at base position 183515
of RP11-1289C17.

Polymorphisms exist between AC110762, AC111001 and AC107223. Data
from AC110762 and AC111001 was used to finish AC107223.
FEATURES
Source
1.183515
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/chromosome="4"
/map="4"
/clone="RP11-1289C17"
/clone_id="RPCT-11"
858. .1120
/rpt_family="MaLR"
repeat_region
2166. .2534
/rpt_family="MaLR"
repeat_region
2824. .2853
/rpt_family="AT_rich"
repeat_region
2854. .2975
/rpt_family="L1"
repeat_region
2982. .3060
/rpt_family="L1"
repeat_region
3062. .3225
/rpt_family="L1"
repeat_region
3656. .4069
/rpt_family="L1"
repeat_region
5312. .5540
/rpt_family="MaLR"
repeat_region
5972. .6275
/rpt_family="Alu"
repeat_region
7353. .7377
/rpt_family="AT_rich"
repeat_region
7873. .8247
/rpt_family="MaLR"
repeat_region
8440. .8467
/rpt_family="AT_rich"
repeat_region
8627. .8656
/rpt_family="AT_rich"
repeat_region
9234. .9269
/rpt_family="(TG)n"
repeat_region
11025. .11331
/rpt_family="L2"
repeat_region
12207. .12425
/rpt_family="MERL_type"
repeat_region
12576. .12804
/rpt_family="MIR"
repeat_region
12920. .12940
/rpt_family="(CAAAA)n"
repeat_region
13283. .13312
/rpt_family="AT_rich"
repeat_region
13477. .13513
/rpt_family="(TAAAAA)n"
repeat_region
14224. .14539
/rpt_family="Alu"
repeat_region
14638. .14677
/rpt_family="AT_rich"
repeat_region
14678. .14928
/rpt_family="Alu"
repeat_region
14929. .14967
/rpt_family="AT_rich"
repeat_region
15564. .15670
/rpt_family="MIR"
repeat_region
15835. .15953
/rpt_family="MIR"
repeat_region
16664. .16956
/rpt_family="Alu"
repeat_region
17507. .17529
/rpt_family="AT_rich"
repeat_region
18412. .18647
/rpt_family="MIR"
repeat_region
19122. .19231
/rpt_family="MIR"
repeat_region
19273. .19441
/rpt_family="MIR"
repeat_region
19627. .19726
/rpt_family="Alu"
repeat_region
19848. .19887
/rpt_family="AT_rich"
repeat_region
19894. .20274
/rpt_family="L1"
repeat_region
21512. .21891

```

```

repeat_region /rpt_family="MaLR"
21892. .21962
/rpt_family="(TTTTTA)n"
repeat_region
21963. .22451
/rpt_family="L1"
repeat_region
22452. .22539
/rpt_family="MaLR"
repeat_region
22570. .22612
/rpt_family="AT_rich"
repeat_region
22617. .22683
/rpt_family="(TTTAA)n"
repeat_region
22979. .23222
/rpt_family="L2"
repeat_region
23403. .23872
/rpt_family="L2"
repeat_region
24170. .24191
/rpt_family="AT_rich"
repeat_region
25774. .25917
/rpt_family="MIR"
repeat_region
27006. .27111
/rpt_family="MIR"
repeat_region
27472. .27554
/rpt_family="MIR"
repeat_region
28297. .28426
/rpt_family="MIR"
repeat_region
28448. .28577
/rpt_family="Alu"
repeat_region
28654. .28843
/rpt_family="MIR"
repeat_region
29441. .29474
/rpt_family="MERL_type"
repeat_region
29982. .30231
/rpt_family="MERL_type"
repeat_region
30232. .30542

```

```

Query Match 3.2%: Score 24; DB 9; Length 183515;
Best Local Similarity 100.0%: Pred. No. 0.35;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 731 AATTAATAATTAATTTTATTTTAA 754
|||||
Db 179480 AATTAATAATTAATTTTATTTTAA 179457

```

```

RESULT 8
CGL236287/c 2201 bp DNA linear PLN 29-MAR-2001
LOCUS
DEFINITION
Carphalea glaucescens chloroplast ndhf gene.
ACCESSION
AJ236287
VERSION
AJ236287.1 GI:6102748
KEYWORDS
NADH dehydrogenase subunit F; ndhf gene.
SOURCE
Carphalea glaucescens.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Plastid Carphalea glaucescens
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Gentianales; Rubiaceae; Rubioidae;
Spermatocoeae; Carphalea.
REFERENCE
1 (bases 1 to 2201)
Bremer,B., Jansen,R.K., Oxelman,B., Backlund,M., Lantz,H. and
Kim,K.J.
More characters or more taxa for a robust phylogeny--case study
from the coffee family (Rubiaceae)
JOURNAL
Syst. Biol. 48 (3), 413-435 (1999)
MEDLINE
22061992
PUBMED
12066290
REFERENCE
2 (bases 1 to 2201)
Bremer,B.
Direct Submission
JOURNAL
Submitted (18-DEC-1998) Bremer B., Department of Systematic Botany,
Uppsala University Villav. 6, S-753 36 Uppsala, SWEDEN
location/Qualifiers
1. .2201
/organism="Carphalea glaucescens"
FEATURES
source

```



```

* 7008 8039: contig of 1032 bp in length
* 8040 8139: gap of unknown length
* 8140 9451: contig of 1312 bp in length
* 9452 9551: gap of unknown length
* 10914 10914: contig of 1363 bp in length
* 10915 11014: gap of unknown length
* 11015 12260: contig of 1246 bp in length
* 12261 12360: gap of unknown length
* 12361 13515: contig of 1155 bp in length
* 13516 13615: gap of unknown length
* 13616 14668: contig of 1053 bp in length
* 14669 14768: gap of unknown length
* 16180 16180: contig of 1412 bp in length
* 16281 16280: gap of unknown length
* 16281 17589: contig of 1309 bp in length
* 17580 17683: gap of unknown length
* 17680 18836: contig of 1147 bp in length
* 18837 18936: gap of unknown length
* 18937 20190: contig of 1254 bp in length
* 20191 20290: gap of unknown length
* 20291 21840: contig of 1550 bp in length
* 21841 21940: gap of unknown length
* 21941 23202: contig of 1262 bp in length
* 23203 23302: gap of unknown length
* 23303 24362: contig of 1060 bp in length
* 24363 24462: gap of unknown length
* 24463 25895: contig of 1433 bp in length
* 25886 25995: gap of unknown length
* 25996 27177: contig of 1182 bp in length
* 27178 28696: contig of 1419 bp in length
* 28697 28797: gap of unknown length
* 28797 30378: contig of 1582 bp in length
* 30379 30478: gap of unknown length
* 30479 31855: contig of 1377 bp in length
* 31856 31955: gap of unknown length
* 31956 33141: contig of 1186 bp in length
* 33142 33241: gap of unknown length
* 33242 34423: contig of 1182 bp in length
* 34424 34523: gap of unknown length
* 34524 36638: contig of 2115 bp in length
* 36639 36738: gap of unknown length
* 36739 37784: contig of 1046 bp in length
* 37785 37884: gap of unknown length
* 37885 39305: contig of 1421 bp in length
* 39306 39405: gap of unknown length
* 39406 40903: contig of 1498 bp in length
* 40904 41003: gap of unknown length
* 41004 42118: contig of 1115 bp in length
* 42119 42218: gap of unknown length
* 42219 43416: contig of 1198 bp in length
* 43417 43516: gap of unknown length
* 43517 45350: contig of 1834 bp in length
* 45351 45450: gap of unknown length
* 45451 47159: contig of 1709 bp in length
* 47160 47259: gap of unknown length
* 47260 48641: contig of 1382 bp in length
* 48642 48741: gap of unknown length
* 48742 50087: contig of 1346 bp in length
* 50088 50187: gap of unknown length
* 50188 52135: contig of 1948 bp in length
* 52136 52235: gap of unknown length
* 52236 54710: contig of 2475 bp in length
* 54711 54810: gap of unknown length
* 54811 55897: contig of 1087 bp in length
* 55898 55997: gap of unknown length
* 55998 57628: contig of 1631 bp in length
* 57629 57728: gap of unknown length
* 57729 59568: contig of 1840 bp in length
* 59569 59668: gap of unknown length
* 59669 62336: contig of 2668 bp in length
* 62337 62437: gap of unknown length
* 62437 64086: contig of 1650 bp in length

```

```

* 64087 64186: gap of unknown length
* 64187 66576: contig of 2390 bp in length
* 66577 66676: gap of unknown length
* 66677 68642: contig of 1966 bp in length
* 68643 68742: gap of unknown length
* 68743 70703: contig of 1961 bp in length
* 70704 70803: gap of unknown length
* 70804 73370: contig of 2567 bp in length
* 73371 73470: gap of unknown length
* 73471 75551: contig of 2181 bp in length
* 75552 75652: gap of unknown length
* 75653 78105: contig of 2353 bp in length
* 78106 78205: gap of unknown length
* 78206 81006: contig of 2802 bp in length
* 81007 81107: gap of unknown length
* 81108 83508: contig of 2402 bp in length
* 83509 83608: gap of unknown length
* 83609 86478: contig of 2770 bp in length
* 86479 88803: gap of unknown length
* 88804 88803: contig of 2325 bp in length

Query Match      3.0%: Score 23; DB 2; Length 122517;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 735 AAATTAATTTTATTTTATTT 757
Db 81591 AAATTAATTTTATTTTATTT 81569

RESULT 10
AC129204
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BB0230P05
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160093 bases at least Q40
Consensus quality: 160951 bases at least Q30
Consensus quality: 160951 bases at least Q20
Insert size: 156000; agarose-fp
Insert size: 163031; sum-of-contrigs
Quality coverage: 7.02 in Q20 bases; agarose-fp
Quality coverage: 6.62 in Q20 bases; sum-of-contrigs

```

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1498: contig of 1498 bp in length
* 1499 1598: gap of unknown length
* 1599 3991: contig of 2393 bp in length
* 3992 4091: gap of unknown length
* 4092 6851: contig of 2760 bp in length
* 6852 14699: gap of unknown length
* 14699 14700: contig of 7748 bp in length
* 14700 14800: gap of unknown length
* 14800 37409: contig of 22609 bp in length
* 37409 37509: gap of unknown length
* 37509 44231: contig of 6723 bp in length
* 44232 44331: gap of unknown length
* 44332 56271: contig of 11940 bp in length
* 56272 56371: gap of unknown length
* 56372 66226: contig of 9855 bp in length
* 66227 66326: gap of unknown length
* 66327 87437: contig of 21111 bp in length
* 87438 87537: gap of unknown length
* 87538 110102: contig of 22565 bp in length
* 110103 137230: gap of unknown length
* 137231 137331: contig of 27028 bp in length
* 137331 164131: gap of unknown length
* 164131 164131: contig of 26801 bp in length.
  
```

FEATURES

```

source
1. 164131
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /chromosome="DNK"
   /clone="RP24-230P5"
   1. 1498
misc_feature
   /note="assembly_name:Contig20"
   1599..3991
misc_feature
   /note="assembly_name:Contig21"
   4092..6851
misc_feature
   /note="assembly_name:Contig22"
   6952..14699
misc_feature
   /note="assembly_name:Contig23"
   clone_end:SP6
   vector_side:right"
   14800..37408
misc_feature
   /note="assembly_name:Contig30"
   clone_end:T7
   vector_side:left"
   37509..44231
misc_feature
   /note="assembly_name:Contig24"
   44332..56271
misc_feature
   /note="assembly_name:Contig25"
   56372..66226
misc_feature
   /note="assembly_name:Contig26"
   66327..87437
misc_feature
   /note="assembly_name:Contig27"
   87538..110102
misc_feature
   /note="assembly_name:Contig28"
   110203..137230
misc_feature
   /note="assembly_name:Contig29"
   137331..164131
misc_feature
   /note="assembly_name:Contig31"
BASE COUNT 50744 a 29926 c 29284 g 53049 t 1128 others
ORIGIN
  
```

Query Match 3.0%: Score 23; DB 2: Length 164131;
 Best Local Similarity 100.0%: Pred. No. 1.2;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 735 AAAAAAAAAATTTTATTTTATTT 757
 Db 131295 AAAAAAAAAATTTTATTTTATTT 131317

RESULT 11
 AC129100
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-1f9, *** SEQUENCING IN PROGRESS ***
 66 unordered pieces.
 AC129100.2 GI:22038210
 HTG: HTGS-PHASE1.
 SOURCE
 Norway rat.
 Rattus norvegicus
 Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;
 Rattus.

REFERENCE

1 (bases 1 to 196029)
 Munzly,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amaralungu,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Boule,S., Brievara,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Kocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunatane,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kover,C.,
 Krtovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M., Louisedge,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maneshwar,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., McWhiney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokkenko,S., Ogub,M., Okunonu,G.,
 Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Qules,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shooashari,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Swalek,A., Taber,P., Tameris,A., Tameris,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telitrod,B., Thomas,N., Thomas,S.,
 Usmanli,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,K.R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 196029)

AUTHORS

Worley,K.C.

TITLE

Direct Submission

JOURNAL

Submitted (27-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE

Submitted (02-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

AUTHORS

Worley,K.C.

TITLE

Direct Submission

JOURNAL

Submitted (02-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Aug 1, 2002 this sequence version replaced g1:21998866.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUVAV
Center clone name: CH230-1F9
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 138974 bases at least Q40
Consensus quality: 148046 bases at least Q30
Consensus quality: 154178 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1171: contig of 1171 bp in length
* 1172 1271: gap of unknown length
* 1272 2975: contig of 1704 bp in length
* 3076 3075: gap of unknown length
* 3076 4476: contig of 1401 bp in length
* 4477 4576: gap of unknown length
* 4577 5645: contig of 1069 bp in length
* 5646 5745: gap of unknown length
* 5746 7172: contig of 1427 bp in length
* 7173 7272: gap of unknown length
* 7273 8427: contig of 1155 bp in length
* 8428 8527: gap of unknown length
* 8528 9597: contig of 1070 bp in length
* 9598 9697: gap of unknown length
* 9698 10849: contig of 1152 bp in length
* 10850 12093: contig of 1144 bp in length
* 12094 12193: gap of unknown length
* 12194 13739: contig of 1546 bp in length
* 13740 13839: gap of unknown length
* 13840 14940: contig of 1101 bp in length
* 14941 15040: gap of unknown length
* 15041 16627: contig of 1587 bp in length
* 16628 16727: gap of unknown length
* 16728 17926: contig of 1199 bp in length
* 17927 18026: gap of unknown length
* 18027 19591: contig of 1565 bp in length
* 19592 21117: gap of unknown length
* 21118 21217: contig of 1426 bp in length
* 21218 23072: gap of unknown length
* 23073 23172: gap of unknown length
* 23173 25701: contig of 2529 bp in length
* 25702 25801: gap of unknown length
* 25802 27506: contig of 1705 bp in length
* 27507 27606: gap of unknown length
* 27607 28822: contig of 1216 bp in length
* 28823 28922: gap of unknown length
* 28923 30312: contig of 1390 bp in length
* 30313 30412: gap of unknown length
* 30413 31772: contig of 1360 bp in length
* 31773 31872: gap of unknown length
* 31873 33897: contig of 2025 bp in length
* 33898 34998: contig of 1001 bp in length
* 34999 35098: gap of unknown length

35099 36734: contig of 1636 bp in length
* 36735 36834: gap of unknown length
* 36835 38609: contig of 1775 bp in length
* 38610 38709: gap of unknown length
* 38710 40577: contig of 1868 bp in length
* 40578 40677: gap of unknown length
* 40678 43341: contig of 2664 bp in length
* 43342 43441: gap of unknown length
* 43442 45264: contig of 1823 bp in length
* 45265 45364: gap of unknown length
* 45365 47920: contig of 2556 bp in length
* 47921 48020: gap of unknown length
* 48021 51219: contig of 3199 bp in length
* 51220 51319: gap of unknown length
* 51320 53170: contig of 1851 bp in length
* 53171 53270: gap of unknown length
* 53271 55133: contig of 1863 bp in length
* 55134 55234: gap of unknown length
* 55235 56653: contig of 1420 bp in length
* 56654 56753: gap of unknown length
* 56754 57980: contig of 1227 bp in length
* 57981 58080: gap of unknown length
* 58081 59357: contig of 1277 bp in length
* 59358 59457: gap of unknown length
* 59458 61465: contig of 2008 bp in length
* 61466 61565: gap of unknown length
* 61566 64484: contig of 2919 bp in length
* 64485 64584: gap of unknown length
* 64585 67652: contig of 3068 bp in length
* 67653 67752: gap of unknown length
* 67753 71133: contig of 3381 bp in length
* 71134 71233: gap of unknown length
* 71234 73475: contig of 2242 bp in length
* 73476 73575: gap of unknown length
* 73576 76247: contig of 2672 bp in length
* 76248 76348: gap of unknown length
* 76349 79097: contig of 2749 bp in length
* 79098 79197: gap of unknown length
* 79198 83156: contig of 3960 bp in length
* 83157 83256: gap of unknown length
* 83257 86344: contig of 3088 bp in length
* 86345 86444: gap of unknown length
* 86445 88999: contig of 2555 bp in length
* 89000 89099: gap of unknown length
* 89100 92534: contig of 3435 bp in length
* 92535 92634: gap of unknown length
* 92635 96128: contig of 3494 bp in length
* 96129 96228: gap of unknown length
* 96229 99739: contig of 3511 bp in length
* 99740 99839: gap of unknown length
* 99840 103358: contig of 3519 bp in length
* 103359 103458: gap of unknown length
* 103459 108285: contig of 4827 bp in length
* 108286 108385: gap of unknown length
* 108386 111692: contig of 3307 bp in length
* 111693 111792: gap of unknown length
* 111793 115975: contig of 4183 bp in length
* 115976 116075: gap of unknown length
* 116076 119975: contig of 3900 bp in length
* 119976 120075: gap of unknown length
* 120076 128264: contig of 8189 bp in length

Query Match 3.0%: Score 23; DB 2; Length 196029;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 729 ACAATAAATTAATTTTATTT 751
|||||

Db 141514 ACAATAAATTAATTTTATTT 141536

RESULT 12
AC122636

LOCUS AC122636 208590 bp DNA linear HTG 14-JUL-2002
 DEFINITION Rattus norvegicus clone CH230-202J22, *** SEQUENCING IN PROGRESS
 AC122636
 AC122636.2 GI:21731454
 HTG: HTGS_PHASE1.
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
 1 (bases 1 to 208590)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsprouck,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbieri,A., Benton,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.T., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homst,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwkw,S., Oguh,M., Okunou,G., Oragunye,N., Oyedero,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojals,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherf,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabar,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 208590)
 AUTHOR Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 208590)
 REFERENCE
 TITLE Direct Submission
 AUTHOR Worley,K.C.
 TITLE Submitted (14-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2002 this sequence version replaced gi:21205953.
 COMMENT
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

Center project name: GKQ
 Center clone name: CH230-202J22
 ----- Summary Statistics
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 192895 bases at least Q40
 Consensus quality: 195064 bases at least Q30
 Consensus quality: 196823 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1026	contig of 1026 bp in length
1127	1126	gap of unknown length
1127	2227	contig of 1101 bp in length
2228	2327	gap of unknown length
2328	3483	contig of 1156 bp in length
3484	3583	gap of unknown length
3584	5729	contig of 2146 bp in length
5730	5829	gap of unknown length
5830	7095	contig of 1266 bp in length
7096	7195	gap of unknown length
7196	8426	contig of 1231 bp in length
8427	8526	gap of unknown length
8527	10655	contig of 2129 bp in length
10656	10755	gap of unknown length
10756	13727	contig of 2972 bp in length
13728	13827	gap of unknown length
13828	18553	contig of 4726 bp in length
18554	18653	gap of unknown length
18654	22873	contig of 4220 bp in length
22874	22973	gap of unknown length
22974	28167	contig of 5194 bp in length
28168	28267	gap of unknown length
28268	37153	contig of 8866 bp in length
37154	37253	gap of unknown length
37254	44057	contig of 6804 bp in length
44058	44157	gap of unknown length
44158	52704	contig of 8547 bp in length
52705	52804	gap of unknown length
52805	67476	contig of 14672 bp in length
67477	67576	gap of unknown length
67577	81856	contig of 14280 bp in length
81857	81956	gap of unknown length
81957	99334	contig of 17378 bp in length
99335	99434	gap of unknown length
99435	121154	contig of 21720 bp in length
121155	121254	gap of unknown length
121255	146708	contig of 25454 bp in length
146709	146808	gap of unknown length
146809	175660	contig of 28852 bp in length
175661	175760	gap of unknown length
175761	208590	contig of 32830 bp in length.

FEATURES
 source 1..208590
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-202J22"

BASE COUNT 60704 a 42199 c 42063 g 60607 t 3017 others
 ORIGIN
 Query Match 3 0%: Score 23: DB 2: Length 208590;
 Best Local Similarity 100.0%: Pred. No. 1.1;
 Matches 23: Conservative 0: Mismatches 0: Gaps 0;


```

OY 729 ACAATAAATTTTATTT 751
|||||
Db 149372 ACAATAAATTTTATTT 149394

RESULT 13
AX346256 11473 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 1327 from Patent WO0200928.
ACCESSION AX346256
VERSION AX346256.1 GI:18494142
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1327 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
Source location/Qualifiers
1..11473
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2662 a 260 c 2742 g 5609 t
ORIGIN

Query Match 2.9%; Score 22; DB 6; Length 11473;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 736 AATAAATTTTATTTATTT 757
|||||
Db 7672 AATAAATTTTATTTATTT 7693

RESULT 14
AX348416 11473 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 111 from Patent WO0202806.
ACCESSION AX348416
VERSION AX348416.1 GI:18614452
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Method and nucleic acids for pharmacogenomic methylation analysis
JOURNAL Patent: WO 0202806-A 111 10-JAN-2002;
EpiGenomics AG (DE)
FEATURES
Source location/Qualifiers
1..11473
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2662 a 260 c 2742 g 5609 t
ORIGIN

Query Match 2.9%; Score 22; DB 6; Length 11473;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 736 AATAAATTTTATTTATTT 757
|||||
Db 7672 AATAAATTTTATTTATTT 7693

RESULT 15
U82212 46387 bp DNA linear HTG 30-AUG-2001
LOCUS Homo sapiens chromosome 10 clone LA10NC01_23_C_3 map
DEFINITION

```

```

ACCESSION 10q26.1-10q26.2, 1 ordered piece.
VERSION U82212
HTG: HTGS_PHASE2; HTGS_CANCELLED.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL 1 (bases 1 to 46387)
REFERENCE 2
AUTHORS Smith, D.R.
TITLE Sequencing of Human Chromosome 10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 46387)
AUTHORS Du, L. and Smith, D.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1996) Bioinformatics Division, Genome
Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
On Dec 9, 1998 this sequence version replaced gi:1773050.
Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
Technology. Data may contain low quality sequence and BAC/Cosmid
vector sequences.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 46387: contig of 46387 bp in length.
location/Qualifiers
1..46387
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q26.1-10q26.2"
/clone="LA10NC01_23_C_3"
BASE COUNT 14232 a 11463 c 10730 g 9962 t
ORIGIN

Query Match 2.9%; Score 22; DB 2; Length 46387;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 738 TAAATTTTATTTATTTCT 759
|||||
Db 23018 TAAATTTTATTTATTTCT 22997

RESULT 16
AL589910 66591 bp DNA linear PRI 12-MAY-2001
LOCUS Human DNA sequence from clone RP11-57L9 on chromosome 6, complete
DEFINITION sequence.
ACCESSION AL589910
VERSION AL589910.9 GI:14041767
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL 1 (bases 1 to 66591)
REFERENCE 2
AUTHORS Garner, P.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On May 15, 2001 this sequence version replaced gi:13897415.
During sequence assembly data is compared from overlapping clones.
where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission

```

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Ch6>
RP11-5719 is from the library RPCT-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-5719 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-5719 is at 1 in this sequence. The true left end of clone RP1-276911 is at 66492 in this sequence. The true right end of clone RP11-358H7 is at 63055 in this sequence.

FEATURES

source

Location/Qualifiers

1..66591

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone_lib="RPCT-11.1"

1..5907

/note="LIPAS repeat: matches 221..6141 of consensus"

repeat_region

5917..6285

/note="LIPAS repeat: matches 5770..6141 of consensus"

repeat_region

6305..6676

/note="MER57B repeat: matches 1..375 of consensus"

repeat_region

6702..7018

/note="AluSg repeat: matches 1..310 of consensus"

repeat_region

7031..7178

/note="74 copies 2 mer aa 58% conserved"

repeat_region

7389..7984

/note="HAL1 repeat: matches 974..1596 of consensus"

repeat_region

8055..8403

/note="LIME repeat: matches 5320..5642 of consensus"

repeat_region

8492..8788

/note="AluSx repeat: matches 1..302 of consensus"

repeat_region

9142..9841

/note="LIMC repeat: matches 2426..2797 of consensus"

repeat_region

9856..10159

/note="AluSx repeat: matches 1..305 of consensus"

repeat_region

10163..11133

/note="LIM4 repeat: matches 2888..3887 of consensus"

repeat_region

13617..13680

/note="MIR repeat: matches 57..125 of consensus"

repeat_region

15298..15401

/note="L2 repeat: matches 2614..2709 of consensus"

repeat_region

15613..15733

/note="LIPB3 repeat: matches 6032..6148 of consensus"

repeat_region

24017..24085

/note="MER58 repeat: matches 129..2279 of consensus"

repeat_region

24308..24478

/note="MIR repeat: matches 12..180 of consensus"

repeat_region

25452..25595

/note="L2 repeat: matches 723..858 of consensus"

repeat_region

26280..26578

/note="AluSg repeat: matches 2..299 of consensus"

repeat_region

27953..28368

/note="L2 repeat: matches 970..1401 of consensus"

repeat_region

28463..29012

/note="MER68A repeat: matches 7..561 of consensus"

repeat_region

30366..30553

/note="MER3 repeat: matches 1..198 of consensus"

repeat_region

32811..33110

/note="AluSg repeat: matches 1..295 of consensus"

repeat_region

33112..34337

/note="LIPAF repeat: matches 4857..6143 of consensus"

repeat_region

34446..34578

/note="AluSg/x repeat: matches 1..133 of consensus"

repeat_region

36021..36191

/note="MIR repeat: matches 58..235 of consensus"

repeat_region

39013..39153

/note="L2 repeat: matches 2602..2745 of consensus"

repeat_region

39742..40051

/note="AluB repeat: matches 1..309 of consensus"

repeat_region

40227..40541

/note="AluSg repeat: matches 1..313 of consensus"

repeat_region

41333..41424

/note="AluYb8 repeat: matches 220..311 of consensus"

repeat_region

42401..42686

/note="AluYb repeat: matches 1..285 of consensus"

repeat_region

42751..43037

/note="LIMB8 repeat: matches 5995..6285 of consensus"

repeat_region

44097..44391

/note="AluSc repeat: matches 1..296 of consensus"

repeat_region

44393..44544

/note="LIPAB repeat: matches 5481..5628 of consensus"

repeat_region

44545..45006

/note="LIPAB repeat: matches 5684..6156 of consensus"

repeat_region

45046..45142

/note="AluYb/FRAM repeat: matches 200..292 of consensus"

repeat_region

45143..45684

/note="MER41D repeat: matches 3..557 of consensus"

repeat_region

45685..45750

/note="FRAM repeat: matches -2..64 of consensus"

repeat_region

45755..45814

/note="30 copies 2 mer gt 10% conserved"

repeat_region

45816..45919

/note="FLAM_C repeat: matches 1..120 of consensus"

repeat_region

45924..46113

/note="AluSg/x repeat: matches 120..301 of consensus"

repeat_region

46763..47089

/note="MIRID repeat: matches 1..338 of consensus"

repeat_region

47534..47759

/note="MIR repeat: matches 33..262 of consensus"

repeat_region

47891..47915

/note="MIR repeat: matches 31..161 of consensus"

repeat_region

51338..51461

/note="62 copies 2 mer at 62% conserved"

repeat_region

52124..52189

/note="MIRID repeat: matches 71..148 of consensus"

repeat_region

52843..52894

/note="26 copies 2 mer aa 76% conserved"

repeat_region

53596..53856

/note="MER6 repeat: matches 255..779 of consensus"

repeat_region

53860..54254

/note="L2 repeat: matches 2061..2450 of consensus"

repeat_region

56309..56598

/note="AluSc repeat: matches 1..290 of consensus"

repeat_region

57577..57877

/note="AluSx repeat: matches 1..299 of consensus"

repeat_region

57982..58720

repeat_region /note="L1p3 repeat: matches 4697. .5442 of consensus"
58908. .59146
/note="L1M5A repeat: matches 6042. .6295 of consensus"
repeat_region 59738. .59925
/note="L1P9A16 repeat: matches 5959. .6145 of consensus"
repeat_region 60924. .61232
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 62016. .62330
/note="AluSg repeat: matches 1. .312 of consensus"
repeat_region 62698. .62982
/note="AluDb repeat: matches 1. .310 of consensus"
repeat_region 63261. .63450
/note="MER5B repeat: matches 8. .178 of consensus"
repeat_region 63844. .64168
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 64825. .65293
/note="L1M4 repeat: matches 2974. .3487 of consensus"
repeat_region 66434. .66591
/note="AluSg repeat: matches 1. .158 of consensus"
BASE COUNT 24499 a 12448 c 11319 g 18325 t
ORIGIN
Query Match 2.9% Score 22; DB 9; Length 66591;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 736 AATAAATTTTATTTATTT 757
|||||
Db 36896 AATAAATTTTATTTATTT 36875
RESULT 17
AC068603/c 97451 bp DNA Linear HTG 02-JUN-2000
LOCUS Homo sapiens chromosome 15 clone CTD-2529A19 map 15q21, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC068603
VERSION AC068603.2 GI:8190151
KEYWORDS HTG: HTGS-PHASED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 97451)
AUTHORS Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A.,
Madan,A., Nesbitt,R., Shaffer,T. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 97451)
AUTHORS Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A.,
Madan,A., Nesbitt,R., Shaffer,T. and Hood,L.
Direct Submission
Submitted (05-MAY-2000) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Jun 2, 2000 this sequence version replaced gi:7708805.
----- Genome Center
Center: Multimegabase Sequencing Center
Genetic code: UMSC
Web site: http://chroma.mbl.washington.edu/msg_mw
Contact: leerowensystemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-Primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

* NOTE: This record contains 96 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
913: contig of 913 bp in length
914 1013: gap of unknown length
1014 1923: contig of 910 bp in length
1924 2023: gap of unknown length
2024 2932: contig of 909 bp in length
2933 3032: gap of unknown length
3033 3948: contig of 916 bp in length
3949 4048: gap of unknown length
4049 4962: contig of 914 bp in length
4963 5062: gap of unknown length
5063 5963: contig of 901 bp in length
5964 6063: gap of unknown length
6064 6971: contig of 908 bp in length
6972 7071: gap of unknown length
7072 7992: contig of 921 bp in length
7993 8092: gap of unknown length
8093 9028: contig of 936 bp in length
9029 9128: gap of unknown length
9129 10060: contig of 932 bp in length
10061 10160: gap of unknown length
10161 11078: contig of 918 bp in length
11079 11178: gap of unknown length
11179 12101: contig of 923 bp in length
12102 13122: contig of 921 bp in length
13123 13222: gap of unknown length
13223 14132: contig of 910 bp in length
14133 14232: gap of unknown length
14233 15142: contig of 910 bp in length
15143 15242: gap of unknown length
15243 16152: contig of 910 bp in length
16153 16252: gap of unknown length
16253 17171: contig of 919 bp in length
17172 17271: gap of unknown length
17272 18178: contig of 907 bp in length
18179 18278: gap of unknown length
18279 19187: contig of 909 bp in length
19188 19287: gap of unknown length
19288 20204: contig of 917 bp in length
20205 20304: gap of unknown length
20305 21232: contig of 928 bp in length
21233 21332: gap of unknown length
21333 22244: contig of 912 bp in length
22245 22344: gap of unknown length
22345 23249: contig of 905 bp in length
23249 23349: gap of unknown length
23350 24263: contig of 914 bp in length
24263 24363: gap of unknown length
24364 25273: contig of 910 bp in length
25274 25373: gap of unknown length
25374 26282: contig of 909 bp in length
26283 26382: gap of unknown length
26383 27302: contig of 920 bp in length
27303 27402: gap of unknown length
27403 28311: contig of 909 bp in length
28312 28411: gap of unknown length
28412 29326: contig of 915 bp in length
29327 29426: gap of unknown length
29427 30322: contig of 896 bp in length
30323 30422: gap of unknown length
30423 31360: contig of 938 bp in length
31361 31460: gap of unknown length
31461 32369: contig of 909 bp in length
32370 32469: gap of unknown length
32470 33391: contig of 922 bp in length
33392 33491: gap of unknown length

```

* 33492 34403: contig of 912 bp in length
* 34404 34503: gap of unknown length
* 34504 35417: contig of 914 bp in length
* 35418 35517: gap of unknown length
* 35518 36427: contig of 910 bp in length
* 36428 36527: gap of unknown length
* 36528 37438: contig of 911 bp in length
* 37439 37538: gap of unknown length
* 37539 38454: contig of 916 bp in length
* 38455 38554: gap of unknown length
* 38555 39480: contig of 926 bp in length
* 39481 39580: gap of unknown length
* 39581 40490: contig of 910 bp in length
* 40491 40590: gap of unknown length
* 40591 41487: contig of 897 bp in length
* 41488 41587: gap of unknown length
* 41588 42503: contig of 916 bp in length
* 42504 42603: gap of unknown length
* 42604 43514: contig of 911 bp in length
* 43515 43614: gap of unknown length
* 43615 44531: contig of 917 bp in length
* 44532 44631: gap of unknown length
* 44632 45539: contig of 908 bp in length
* 45540 45639: gap of unknown length
* 45640 46569: contig of 930 bp in length
* 46570 46669: gap of unknown length
* 46670 47575: contig of 906 bp in length
* 47576 47675: gap of unknown length
* 47676 48596: contig of 921 bp in length
* 48597 48696: gap of unknown length
* 48697 49631: contig of 935 bp in length
* 49632 49731: gap of unknown length
* 49732 50658: contig of 927 bp in length
* 50659 51712: gap of unknown length
* 51713 51812: contig of 954 bp in length
* 51813 52724: contig of 912 bp in length
* 52725 52824: gap of unknown length
* 52826 53748: contig of 924 bp in length
* 53749 53848: gap of unknown length
* 53849 54761: contig of 913 bp in length
* 54762 54861: gap of unknown length
* 54862 55790: contig of 929 bp in length
* 55791 55890: gap of unknown length
* 55891 56797: contig of 907 bp in length
* 56798 56897: gap of unknown length
* 56898 57814: contig of 917 bp in length
* 57815 57914: gap of unknown length
* 57915 58827: contig of 913 bp in length
* 58828 58927: gap of unknown length
* 58929 59842: contig of 915 bp in length
* 59843 59942: gap of unknown length
* 59944 60862: contig of 920 bp in length
* 60863 60962: gap of unknown length
* 60963 61877: contig of 915 bp in length
* 61878 61977: gap of unknown length
* 61979 62898: contig of 921 bp in length
* 62899 62998: gap of unknown length
* 62999 63961: contig of 963 bp in length
* 63962 64061: gap of unknown length
* 64062 64998: contig of 937 bp in length
* 64999 65098: gap of unknown length
* 65099 66014: contig of 916 bp in length
* 66015 66114: gap of unknown length
* 66115 67030: contig of 916 bp in length
* 67031 67130: gap of unknown length
* 67131 68038: contig of 908 bp in length
* 68039 68138: gap of unknown length
* 68139 69055: contig of 917 bp in length
* 69056 69155: gap of unknown length
* 69156 70088: contig of 933 bp in length
* 70089 70188: gap of unknown length
* 70189 71108: contig of 920 bp in length

```

```

* 71109 71208: gap of unknown length
* 71209 72126: contig of 918 bp in length
* 72127 72226: gap of unknown length
* 72227 73151: contig of 925 bp in length
* 73152 73251: gap of unknown length
* 73252 74165: contig of 914 bp in length
* 74166 74265: gap of unknown length
* 74266 75197: contig of 932 bp in length
* 75198 75297: gap of unknown length
* 75298 76205: contig of 908 bp in length
* 76206 76305: gap of unknown length
* 76306 77199: contig of 894 bp in length
* 77200 77299: gap of unknown length
* 77300 78310: contig of 911 bp in length

```

Query Match 2.9% Score 22; DB 2; Length 97451;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 736 AATAAATTTTATTTTATTT 757
|||||

Db 31336 AATAAATTTTATTTTATTT 31315

RESULT 18
AC004022 116552 bp DNA linear PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone GSI-155M11 from 7q21-q22, complete sequence.
DEFINITION AC004022.1 GI:3598727
ACCESSION AC004022.1
VERSION AC004022.1
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 116552)
Hinds, R., Tin-Wollam, A., Becker, M. and Stromwall, C.
JOURNAL The sequence of Homo sapiens BAC clone GSI-155M11
Unpublished

REFERENCE
AUTHORS 2 (bases 1 to 116552)
Waterston, R.

JOURNAL Direct Submission
TITLE Submitted (30-JUL-1998) Department of Genetics, Washington
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 116552)

REFERENCE
AUTHORS 4 (bases 1 to 116552)
Waterston, R.

JOURNAL Direct Submission
TITLE Submitted (12-SEP-1998) Department of Genetics, Washington
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 116552)

REFERENCE
AUTHORS 5 (bases 1 to 116552)
Waterston, R.

JOURNAL Direct Submission
TITLE Submitted (21-DEC-1999) Department of Genetics, Washington
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 13, 1998 this sequence version replaced g1:2798822.

COMMENT
----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: saplens@wustl.wustl.edu

----- Summary Statistics

Center project name: H_GSI5M11

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).
Cell line: lymphoblastoid

Haplotypes: two

VECTOR: pBel0BAC

Selection: Chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is GS1-293C5, 200 bp overlap.
Actual start of this clone is at base position 1 of GS1-155M11.
Actual end is at 116552 of GS1-155M11.

Location/Qualifiers

1. 116552

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-q22"

/clone="GS1-155M11"

/clone_1id="GSBAC1"

924. 971

/rpt_family="Alu"

repeat_region

1002. 2267

/rpt_family="L1"

repeat_region

2294. 3244

/rpt_family="L1"

repeat_region

3271. 3468

/rpt_family="L1"

repeat_region

3628. 3795

/rpt_family="L1"

repeat_region

5164. 5461

/rpt_family="Alu"

repeat_region

8216. 8481

/rpt_family="Alu"

repeat_region

8632. 8788

/rpt_family="MIR"

repeat_region

10077. 10178

/rpt_family="MIR"

repeat_region

11276. 11452

/rpt_family="MIR"

gene

<13319. 35070

/gene="WUGSC:H.GS155M11.3"

join(<13319. 35070, 31092, 33303, 33485, 34858, 35070)

/gene="WUGSC:H.GS155M11.3"

/note="Neurabin-like; similar to U72994 (PID:g2623757); H.GS155M11.3"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC35294.1"

/db_xref="GI:3598728"

/translation="LINEAR PETERLIN DSKALUTR ROLSVKRRRORPSRRLVYDSSYSD
GSDSLERKNTFTNDSPSSSSADISGIGAPKTPGTSLSLSDPDLDDOSPKH
SOGONRAVQVMSVQVSHWLMSTLNLEDYVEFSQNTTGECOLLLOLDGNKIKALGMAS
QDRAVVAKKLKEKMSLEKARKAOKEMKORERKLRRKROBOMORRKSRTKMTSTAE
GAGEQ"

repeat_region

14804. 14983

/rpt_family="MER1_type"

repeat_region

19556. 19778

/rpt_family="MIR"

repeat_region

19823. 19892

/rpt_family="MIR"

repeat_region

19893. 20189

/rpt_family="Alu"

repeat_region

20190. 20305

/rpt_family="MIR"

repeat_region

21355. 21440

/rpt_family="U13"

repeat_region

22190. 22493

/rpt_family="Alu"

repeat_region

23027. 23278

/rpt_family="MER1_type"

repeat_region

23590. 23682

/rpt_family="MIR"

repeat_region

26847. 27111

/rpt_family="Alu"

repeat_region

28064. 28202

/rpt_family="MIR"

repeat_region

32387. 32446

/rpt_family="MIR"

repeat_region

32447. 32754

/rpt_family="Alu"

repeat_region

32755. 32847

/rpt_family="MIR"

repeat_region

39036. 39202

/rpt_family="Alu"

repeat_region

39050. 39343

/note="match to EST R42087 (NID:9817025) yf98h04.s1"

misc_feature

39063. 39323

/note="match to EST F04750 (NID:9671439) "

misc_feature

39368. 39921

/note="match to EST R36730 (NID:9793631) y906h04.r9"

misc_feature

39568. 39897

/note="match to EST R22722 (NID:9777479) y908h04.r1"

misc_feature

39815. 39967

/note="match to EST R36730 (NID:9793631) y906h04.r9"

misc_feature

39841. 39983

/note="match to EST R22722 (NID:9777479) y908h04.r1"

misc_feature

40960. 40960

/note="match to EST R42416 (NID:9817182) yf92e01.s1"

misc_feature

41167. 41167

/note="match to EST R42416 (NID:9817182) yf92e01.s1"

repeat_region

42277. 42585

/rpt_family="Alu"

misc_feature

43124. 43437

/note="match to EST R12373 (NID:9765449) yf28b04.s1"

misc_feature

43127. 43527

/note="match to EST T90908 (NID:9722821) yd48g09.s1"

misc_feature

43353. 43478

/note="match to EST R12373 (NID:9765449) yf28b04.s1"

misc_feature

43452. 43827

/note="match to EST AA953386 (NID:93117533) on63g09.s1"

misc_feature

43462. 43610

/note="match to EST T90908 (NID:9722821) yd48g09.s1"

misc_feature

43465. 43871

/note="match to EST R43174 (NID:9825400) y918c05.s1"

misc_feature

43486. 43871

/note="match to EST AA907345 (NID:93042805) oj70e09.s1"

misc_feature

gene

complement(43711. 69242)

/gene="PONI"

/complement(join(43711. 43869, 46872. 47100, 51052. 51133, 52778. 52978, 56218. 56344, 60089. 60257, 61501. 61556, 63090. 63160, 69169. 69242))

/note="match to P27169 (PID:9130675): but note polymorphism at position 55; h_GS155M11.1"

/product="serum paraoxonase[esterase 1]"

/protein_id="AAC35293.1"

/db_xref="GI:2795823"

/translation="MAKLIALTLGLGALFNNHSSVOTRLNAREVQVPELPCNLT
VKGITGSEDLLEIPNLGAFISSGLKYPGKSFNPSPCKILIMDNEEDPTVLELGI
TGSKEFVSSEFNHGISLTFEDNEMAYLLVNNHDKASTVELFKFOEEKSLHLKTR
HKLPLNDINIVAGPEHEFGTDHFLDYPLOSWEKMYGLANSYVYVSPSRVYAE
GFDLPANGINISPDGKYVYIAELLAKHIEYKHAMWTLPLSLIDENTLYNDISDPE
TGDINVGCHPNMKITFFYDSENPSPASEVIRIONILTEEPKTVQVYAEKSTVLOGSTVA
SVYKGLLIGVFNKALYCEL"

Query Match 2.9%; Score 22; DB 9; Length 116552;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 AATAAATAAATTTTATTTT 752
|||||
Db 105520 AATAAATAAATTTTATTTT 105541

RESULT 19
AL691434 122972 bp DNA linear HTG 24-APR-2002
LOCUS Homo sapiens chromosome 10 clone CTD-2095B6, *** SEQUENCING IN
DEFINITION PROGRESS *** 5 unordered pieces.
ACCESSION AL691434
VERSION AL691434.5 GI:20330345
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVERFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Brown, A.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 25, 2002 this sequence version replaced g1:20152449.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BK2095B6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 98% of reads
Chemistry: Dye-terminator; 1% of reads
Consensus quality: 121473 bases at least Q40
Consensus quality: 122023 bases at least Q30
Consensus quality: 122343 bases at least Q20
Insert size: 122572; sum-of-contrigs
Insert size: 139827; 2.7% error; agarose-fp
Quality coverage: 7.82x in Q20 bases; sum-of-contrigs Quality
coverage: 7.19x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 99034: contrig of 99034 bp in length
* 99035 99134: gap of 100 bp
* 99135 105347: contrig of 6213 bp in length
* 105348 105447: gap of 100 bp
* 105448 110538: contrig of 5091 bp in length
* 110539 110638: gap of 100 bp
* 110639 120626: contrig of 9988 bp in length
* 120627 120726: gap of 100 bp
* 120727 122972: contrig of 2246 bp in length.
Location/Qualifiers

FEATURES

source 1. 122972
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="CTD-2095B6"
/clone_11b="CIT-HSP-D1"
1. 99034
/note="assembly-fragment:00195
fragment_chain:1"
99135. 105347
/note="assembly-fragment:01101
fragment_chain:1"
105448. 110538
/note="assembly-fragment:02472
fragment_chain:1"
110639. 120626
/note="assembly-fragment:03121.0"
120727. 122972
/note="assembly-fragment:03144"
25242 a 31284 c 32711 g 33333 t 402 others
misc_count

Query Match 2.9%; Score 22; DB 2; Length 122972;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 738 TAAATTTTATTTTATTTCT 759
|||||
Db 29548 TAAATTTTATTTTATTTCT 29569

RESULT 20
AL162274 148891 bp DNA linear PRI 03-DEC-2001
LOCUS Human DNA sequence from clone RP11-45A17 on chromosome 10 contains
DEFINITION the BNIP3 gene for BCL2/adenovirus E1B 19kD-interacting protein 3
and Cpg islands, complete sequence.
ACCESSION AL162274
VERSION AL162274.17 GI:12214292
KEYWORDS HTG; adenovirus; BCL2; BNIP3; Cpg island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 148891)
AUTHORS Brown, A.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jan 13, 2001 this sequence version replaced g1:11544962.

* During sequence assembly data is compared from overlapping clones.
* Where differences are found these are annotated as variations
* together with a note of the overlapping clone name. Note that the
* variation annotation may not be found in the sequence submission
* corresponding to the overlapping clone, as we submit sequences with
* only a small overlap as described above.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/c-elegans/wormpep This sequence
was generated from part of bacterial clone contrigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-45A17 is from the library RP01-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-45A17 The true
left end of clone RP11-140A10 is at 124384 in this sequence. This
sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES

source

1. 148891

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-45A17"

/clone_lib="RPC1-11.1"

4. .638

misc_feature

/note="match: GSS: Em:AQ199021"

repeat_region

/note="AluJo repeat: matches 7. .293 of consensus"

repeat_region

/note="AluSx repeat: matches 1. .302 of consensus"

gene

/complement(1627. .15797)

mRNA

/complement(join(1627. .2496,4565. .4714,4791. .4897, 6961. .7045,7720. .7870,15688. .15797))

/gene="BNIP3"

/product="bA45A17.1 (BCL2/adenovirus E1B 19kD-interacting protein 3)"

/note="match: cDNAs: Em:AF002697 Em:U15174 Em:AF243515 Em:AF041054 Em:AF067396 Em:AF067395 Em:AL12665 Em:AB004788 Em:AF079221 Em:M94291 Em:M31178

match: ESTs: Em:BF315208 Em:BE905125 Em:BE619261

Em:BE407953 Em:BE879860 Em:BF127765 Em:BF18024 Em:N28896

Em:BE884155 Em:M15558 Em:AW985532 Em:M15599 Em:BF16555

Em:BE851099 Em:BE689682 Em:BE311006 Em:C89438 Em:AI95667

Em:BE854060 Em:AW968613 Em:BE738320 Em:RI8500 Em:BF021518

Em:N25150 Em:AF202371 Em:A1047881 Em:AW921704 Em:AO600956

Em:BE629204 Em:C15736 Em:BE850955 Em:BE650348 Em:AI084859

Em:AA095273 Em:AA038867 Em:AI098525 Em:AW012733

Em:AA028224 Em:AI098526 Em:AV721096 Em:N23492 Em:BE737768

Em:W08025 Em:BF012169 Em:R34992 Em:H56529 Em:AI528169

Em:AA106511 Em:AA959197 Em:M15973 Em:F07689 Em:AM166682

Em:H35608 Em:H21080 Em:AI950646 Em:BE620872 Em:N77697

Em:AI367455 Em:R26638 Em:AW989885 Em:F07349 Em:BE128625

Em:F07665 Em:AI196455 Em:BE620465 Em:AA063521 Em:AV661296

Em:H24215 Em:HI0866 Em:AW967789 Em:RI9446 Em:AA93817

Em:W76907 Em:W09821 Em:AA035669 Em:AW131053 Em:AI304210

Em:AI183556 Em:H83602 Em:BF244972 Em:BE981805 Em:AW971559

Em:RI3791 Em:AI214880 Em:M85629 Em:AV001095 Em:BE127163

Em:BE972728 Em:AW552830 Em:AW138491 Em:BE126315

Em:AI857795 Em:AW516227 Em:N87221 Em:AW144834 Em:AW214824

Em:AI795904 Em:BF244407 Em:BF244239 Em:N77279 Em:AI343836

Em:AW53554 Em:BE633064 Em:AI768367 Em:BF244888"

/evidence="not_experimental"

complement(1627)

/gene="BNIP3"

complement(1631. .1636)

/gene="BNIP3"

complement(2126. .2261)

/gene="BNIP3"

/note="match: STS: Em:G54089"

complement(join(2451. .2496,4565. .4714,4791. .4897, 6961. .7045,7720. .7870,15688. .15793))

/gene="BNIP3"

/note="match: proteins: SW:Q12983 SW:O55003 SW:O60238 SW:Q9627"

/codon_start=1

/evidence="not_experimental"

/product="bA45A17.1 (BCL2/adenovirus E1B 19kD-interacting protein 3)"

/protein_id="CAD13201.1"

/db_xref="GI:11784058"

/translation="MSONGAPGQGESLQGSVWELHPSNNGNGSPASVSIYNGME KILLDAQHSGRSSSSHCSDSPRSQTPDQTNRASETPTHSIGENSSQSEDDIER RKEVESILKKNSDWMDSRPENIPKPEFLFKHPRRTATILSMRNTSVMKKGIFSAE

FLKVPFLSLLSHLAIGUIGYIGRRLLTSTSTF"

2867. .2914

/note="L2 repeat: matches 2619. .2664 of consensus"

3274. .3365

/note="MER5A repeat: matches 35. .141 of consensus"

3834. .4099

/note="AluSg repeat: matches 1. .277 of consensus"

4380. .4470

/note="L1ME3A repeat: matches 6019. .6119 of consensus"

4564. .4844

/note="match: GSS: Em:A2445644"

6200. .6313

/note="FLAM_C repeat: matches 1. .127 of consensus"

8317. .8606

/note="AluJo repeat: matches 1. .298 of consensus"

8623. .8740

/note="L1MC/D repeat: matches 5318. .5436 of consensus"

10297. .10513

/note="MER21B repeat: matches 43. .360 of consensus"

10849. .10910

/note="MER41A repeat: matches 400. .460 of consensus"

10923. .11212

/note="AluSg repeat: matches 6. .297 of consensus"

11373. .11631

/note="MLT1A1 repeat: matches 56. .348 of consensus"

13528. .13613

/note="AluJo/FRAM repeat: matches 211. .296 of consensus"

14388. .14518

/note="FLAM_C repeat: matches 1. .131 of consensus"

15322. .17021

/note="Cpg island"

/evidence="not_experimental"

15798. .16085

/note="8 copies 36 mer 60% conserved"

15940. .16177

/note="14 copies 17 mer 58% conserved"

15941. .16094

/note="77 copies 2 mer 99 60% conserved"

17123. .17388

/note="AluB repeat: matches 14. .301 of consensus"

17444. .17570

/note="FLAM_C repeat: matches 1. .130 of consensus"

19523. .19730

/note="AluSg repeat: matches 1. .297 of consensus"

20302. .20739

/note="6 copies 73 mer 97% conserved"

20748. .21044

/note="AluSg repeat: matches 1. .299 of consensus"

21170. .21279

/note="L1PA5 repeat: matches 5129. .5238 of consensus"

21282. .21328

/note="Alu repeat: matches 7. .53 of consensus"

21350. .22225

/note="L1PA5 repeat: matches 5254. .6141 of consensus"

22207. .22614

/note="L1ME1 repeat: matches 5468. .5926 of consensus"

22615. .22798

/note="LTR8 repeat: matches 513. .691 of consensus"

22808. .23109

/note="AluSg repeat: matches 1. .302 of consensus"

23300. .23411

/note="FLAM_C repeat: matches 7. .118 of consensus"

23974. .24273

/note="AluJ repeat: matches 1. .300 of consensus"

24274. .24303

/note="15 copies 2 mer ta 93% conserved"

24666. .24961

/note="AluSg repeat: matches 1. .296 of consensus"

24997. .25274

/note="AluJo repeat: matches 1. .297 of consensus"

25278. .25411

/note="L1MC5 repeat: matches 7171. .7276 of consensus"

25412. .25713

repeat_region


```

repeat_region      9521..9814
                    /note="Alusx repeat: matches 6..299 of consensus"
repeat_region      9815..10430
                    /note="LIM4c repeat: matches 986..1595 of consensus"
gene               /note="LIM4c repeat: matches 986..1595 of consensus"
                    /gene="bA189K21.2"
                    /complement(10696..11966)
                    /gene="bA189K21.2"
                    /complement(10696..11966)
CDS                /gene="bA189K21.2"
                    /note="bA189K21.2 (glucosaminyl (N-acetyl) transferase 1,
                    core 2
                    (beta-1,6-N-acetylglucosaminyltransferase)
                    (GCNT1) pseudogene)
                    match: proteins Tr:Q19730 Sw:Q06430 Sw:P97402 Sw:Q09324
                    Tr:Q19729 Tr:Q02234 Tr:Q05395 Tr:Q02314 Tr:Q64165
                    Tr:Q02315 Sw:Q02742 Sw:Q92180"
                    /pseudo
                    /codon_start=1
                    /evidence=not_experimental
repeat_region      12419..12978
                    /note="LI repeat: matches 4752..5332 of consensus"
repeat_region      13024..13288
                    /note="MER46B repeat: matches 92..236 of consensus"
repeat_region      13289..13546
                    /note="LI repeat: matches 4504..4755 of consensus"
repeat_region      13547..13850
                    /note="MER7A repeat: matches 2..345 of consensus"
repeat_region      13851..13926
                    /note="LI repeat: matches 4428..4504 of consensus"
repeat_region      13920..14683
                    /note="LIM4 repeat: matches 2532..3356 of consensus"
repeat_region      14704..15020
                    /note="Alusx repeat: matches 1..312 of consensus"
repeat_region      15035..15728
                    /note="LIM1 repeat: matches 5476..6163 of consensus"
repeat_region      15811..16101
                    /note="Alusg repeat: matches 5..295 of consensus"
repeat_region      16122..16572
                    /note="MTJC repeat: matches 1..466 of consensus"
repeat_region      16595..16906
                    /note="MER1B repeat: matches 1..337 of consensus"
repeat_region      17882..18023
                    /note="FLAMC repeat: matches 1..142 of consensus"
repeat_region      18128..18173
                    /note="23 copies 2 mer tg 87% conserved"
repeat_region      18405..18704
                    /note="Aluyb8 repeat: matches 1..306 of consensus"
repeat_region      18727..19022
                    /note="Alusg repeat: matches 1..296 of consensus"
repeat_region      19963..20167
                    /note="LIMC1 repeat: matches 6122..6325 of consensus"
repeat_region      21069..21381
                    /note="Aluyb repeat: matches 1..312 of consensus"
repeat_region      22822..22961
                    /note="MTJ1 repeat: matches 378..516 of consensus"
repeat_region      23800..24092
                    /note="Alusx repeat: matches 4..297 of consensus"
repeat_region      24378..24465
                    /note="5S repeat: matches 1..90 of consensus"
repeat_region      24517..24626
                    /note="LIM45A repeat: matches 6185..6295 of consensus"
repeat_region      24799..25103
                    /note="Alusg repeat: matches 1..305 of consensus"
repeat_region      25848..26302
                    /note="LIP48 repeat: matches 5709..6163 of consensus"
repeat_region      26316..26368
                    /note="MER41A repeat: matches 401..453 of consensus"
repeat_region      26455..26821
                    /note="LIM2 repeat: matches 947..1327 of consensus"
repeat_region      26846..26997
                    /note="Aluy repeat: matches 148..299 of consensus"
repeat_region      27011..27832
                    /note="LIM2 repeat: matches 1318..2122 of consensus"
repeat_region      27833..28148
                    /note="Alusx repeat: matches 1..312 of consensus"

```

```

repeat_region      28149..28222
                    /note="LIM2 repeat: matches 2122..2193 of consensus"
repeat_region      28223..28328
                    /note="LIM2 repeat: matches 6202..6308 of consensus"
repeat_region      28473..28779
                    /note="Aluy repeat: matches 1..306 of consensus"
repeat_region      28851..29193
                    /note="Aluy repeat: matches 1..306 of consensus"

Query Match      2.9% Score 22 DB 9 Length 151136;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 733 TAAATAAATTTTATTTTA 754
Db 102716 TAAATAAATTTTATTTTA 102695

RESULT 22
AC021043/c 154716 bp DNA linear PLN 29-JUL-2000
LOCUS Arabidopsis thaliana chromosome I BAC F28N24 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC021043
VERSION AC021043.4 GI:7340331
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
1 (bases 1 to 154716)
Fedeerspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Miranda,M., Brooks,S.,
Buehler,E., Chao,Q., Chin,C., Chlou,J., Choi,E., Gonzalez,A.,
Hwang,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M.,
Lenz,C., Liu,A., Liu,S., Mukharsky,N., Pham,P., Sakano,H.,
Shim,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A.
and Davis,R.W.
Unpublished
2 (bases 1 to 154716)
Fedeerspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chlou,J., Choi,E., Dunn,P., Gonzalez,A., Hwang,B., Kim,C.,
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shim,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission
Submitted (12-JAN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 154716)
Fedeerspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chlou,J., Choi,E., Dunn,P., Gonzalez,A., Hwang,B., Kim,C.,
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shim,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission
Submitted (29-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 154716)
Fedeerspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
and Davis,R.W.
Direct Submission
Submitted (27-JUL-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
5 (bases 1 to 154716)
Fedeerspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Altafi, H., Nguyen, M., Lam, B., Southwick, A., Ecker, J., Theologis, A. and Davis, R.W.
Direct Submission
Submitted (29-Jul-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Mar 29, 2000 this sequence version replaced gi:7248938.
Bases 131,240-154,716 of clone F28N24 overlap with bases 1-23,716
of IGF BAC clone Fik23, gb|AC007508.
Location/Qualifiers
1. 154716
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F28N24"
2391..5038
/gene="F28N24.1"
/note="Erd2"
join(2484..2574,2974..3239,3947..4062,4462..4535,
4622..4685,4774..4810)
/gene="F28N24.1"
/note="Identical to gb|A49677"
/codon_start=1
/product="endoplasmic reticulum retention receptor Erd2"
/protein_id="AAF88108.1"
/db_xref="GI:9502409"
/translation="MNIFRAGDMSHLISVLILKITYATKSGASISLKTOLVALVF
LTRYDLFEDVSYLTIKIVFIASLSIAIWMCRMRLPVRSDKLDITPEHOYVL
ACFVGLINKEFTYQEVFMASITLSEAVALLPOLVILORSGVDNLGOIYVFQAY
KGLYIINWIRYFTEDHFTRMIAVSGLVOTALVADFFYYIISMKTNKILPA"
complement(8864..10728)
/gene="F28N24.2"
complement(join(8864..9103,9190..9325,9410..9534,
9620..9974,10084..10311,10385..10728))
/gene="F28N24.3"
/note="member of highly conserved multi-gene family in
arabidopsis, other members include At2g34250, gb|AAC27401
and F9K20.24, gb|AAC83037"
/codon_start=1
/product="Putative protein transport protein SEC61 alpha
subunit"
/protein_id="AAF88109.1"
/db_xref="GI:9502410"
/translation="MGGGFRVLHLVRRPLAPLREYQADRKIPFREKVIYVLSLFI
LYCSQLPIYGIHSTGADPFYRMVILASNRGTMEIGTPIVTSGLVQLAGSKII
EYDNNVREDRALNGAOKLWILSAIGEVAVVLSMGPGVGLGNAIILIILOLF
AGTIVICDELQKXGSGSISLFIATNICSITWKASPTITNGRAEPEGAVIA
LEHMLITKSNKYAALNQAFYRONLPRVNLATVLIPLVITFQGRVVLPRKESAR
GQGSYPIKLFYTSNPIILQSALVSNLFIQLYRKSGNFYVNLGOWMESESG
OSIPVGLAVYLITAPASFADMAHPHLFVIFMLTALFSTKMIIEVSGSSARDA
KOLKEQWMPGRHSNLOKELNRYIPTAAFGVIGALTVLADFMGAISGSGTILL
AVTIYQYEFPEKKEKASELGFPGF"
13891..15270
/gene="F28N24.3"
13891..15270
/gene="F28N24.3"
/note="Unknown protein"
/codon_start=1
/protein_id="AAF88110.1"
/db_xref="GI:9502411"
/translation="METLKKNPSVRKSKLAFQVCNLRITTSRVSSNNGIGICML
KSONPNFDEDDGSDVDLKTSSSRSGEIVYRENRNLVLAQVAKIFASTYSIKAA
YAELOAQRPPYNDALQADATVEELRALSELKRSFELKELNTLSPQVLAIEIOE
OLMRVETITIKKLEPEYERKOLKIDELKMFEEESLYVVKSLKRTSASGSVGDNI
EIRNLNLSFYOVLFPTLRVSVPFKLIYKEMKESWMDAANAASVSNVKNASTVF
ARSHRCFAFESFVCGKMFENFADPFSRREFEKLRSVDPIQYILFRNGSSFARVY
HAYLSVNAKMECSFEGNLNQRKLVNSGGFPPSGFATCEAAKRIWLACIAFSLSG
NVTYOLKGRGCRFSQVYMSKSGDESLFSGDSDIRVGFVYVPGKIGENIOSQVY
LTPVNGFTS"
complement(25923..26801)
/gene="F28N24.4"
complement(join(25923..26566,26666..26801))
/gene="F28N24.4"
CDS

/note="Hypothetical protein"
/codon_start=1
/protein_id="AAF88112.1"
/db_xref="GI:9502413"
/translation="MKRGIDMARSYNDHSSOEQFNGPESNSSTFGMKALISHSPPK
SRRSYKERVVNPMEKSGSRKKGDTTPSSSMARKGQKPIKSPFPRRIYKCSST
KGCPRKQVENSRRDPTWILITYSNHPVLPVLSITNGKPKPEPEPEPEPEPE
AAEEDNKPMVLGRGILETPVSCVDEHAFMTEMETTSTLSEPIFSSEKRTLVAGADV
AVFFPMGEDESLFDELGEPECSVVFHRSSVVSQVEIF"
complement(32369..33038)
/gene="F28N24.5"
complement(join(32369..32623,32955..33038))
/gene="F28N24.5"
/note="Unknown protein"
/codon_start=1
/protein_id="AAF88111.1"
/db_xref="GI:9502412"
/translation="MLKLEIALOREYEDLKITYKESLHOREQKVEKIGEMFOYKIK
SPRGNNVNLLEAFEEGLNVAQAASCLDSFAEALVAPQSKDKLCSYDLDYTLV
KALVPSVPL"
complement(37348..38516)
/gene="F28N24.6"
/note="PEX7"
complement(37514..38467)
/gene="F28N24.6"
/note="CDNA gb|AF130973"
/codon_start=1
/product="peroxisomal targeting signal type 2 receptor,
pex7p"
/protein_id="AAF88113.1"
/db_xref="GI:9502414"
/translation="MPVFRKPFNGYSVRFSPFYESRLAVATAQNGILGNRIHVL
APGAGVTESSYDTPADAVDYDCWSESDSVLIAIGGSVKIYTAALPPSPPIRSF
QEHAREQSVVNPPIRRDPSLTLSWDDTPIILAMAPRPSVTFKHAQVQAOANPK
HCDVYASASGCDTLRIKIVYREBGSIMITPAHDFELSCDMKXIDCIATTSYDVKV
WADVRSYVPLAVLNGHYAAKRVAFSHRSLSLASCSDVSCLDMDYMDVALGRY
DHHTFVAGIDMSVLEGLMASTGMDLVYVWQDMPRAS"
38798..39929
/gene="F28N24.7"
join(38798..38917,39200..39248,39482..39544,39645..39709,
39834..39929)
/gene="F28N24.7"
/note="Unknown protein"
/codon_start=1
/protein_id="AAF88114.1"
/db_xref="GI:9502415"
/translation="MEITTEGVNNMNLAVDQKRRIOVSTTKRPLFFVYNLARKYQ
QITDVELSALGALATVTVVAEILKNGFAVEKKIMTSTVIDIKDSRGRPOKAKIET
TLAKSEKDELMAAANEKKAEOQN"
43556..46544
/gene="F28N24.8"
join(43556..43767,44374..45422,45503..46240,46329..46544)
/gene="F28N24.8"
join(44466..45422,45503..46240,46329..46367)
/gene="F28N24.8"
/note="Unknown protein"
/codon_start=1
/protein_id="AAF88115.1"
/db_xref="GI:9502416"
/translation="MEERLNFSEVPLSTRMOKTAGVSVRRKSNNTDYDSKTS
PSVPVLYVMGLDEVEPASPPTWQDAGRLKGNFQVQVCLMKEEQVFTPL
PGKAVDANMPTLRDSSKGOVEESDDDDVFSADLDLISRDYSFNNSISGSEYG
GVEKTPDAQSRDPMASREFLPAAKAMVEOSHVSNSKPSFMAPEPTIOLRELVPGE
KOQPNRVDVSIPEVYHODIDDEESVGEEDDVCSEYAVLTKRGCGMLPOLCPKDS
LCMLNTVGRFKAKNHSPITPSHDVYKSKAQLKRYRQSKYKLLADSVSKHLSKVY
HSPVHPSGKFKNSNSNLISAKKSSSPYRHSRCSPFRSGNSPLIHAGPEYRRE
TENLANRKLNSMTRNISRISQELVPKSNGSLILEKTVYLDENDMTIDQNSNLMIF
PEEADNRKRPDANDELAFENISIRSEEMVGNELVSISSGDFRSPVAPSPKRPSE
SWLCHNLPSVTSOITSRRVHPNPQKDLTEYRNGTKMETIVKTSYMHRRHIRYSEE
LVAHNSHOSKT"
complement(46968..48530)
/gene="F28N24.9"
complement(46968..48530)
CDS


```
repeat_region /rpt_family="L2"
8845. .9042
/rpt_family="MIR"
repeat_region /rpt_family="AT_rich"
9215. .9255
complement(9400. .9564)
/rpt_family="MIR"
repeat_region complement(10625. .11183)
/rpt_family="L2"
11424. .11729
/rpt_family="AluIo"
repeat_region complement(12401. .12874)
/rpt_family="MLTID"
13237. .13270
/rpt_family="TTTTGn"
repeat_region complement(13271. .13550)
/rpt_family="AluSg"
14205. .14261
/rpt_family="Alu"
repeat_region 14262. .14468
/rpt_family="AluSg/X"
15020. .15047
/rpt_family="(TG)n"
repeat_region complement(15744. .16059)
/rpt_family="LIM4"
repeat_region complement(17100. .17412)
/rpt_family="AluSg"
17938. .18029
/rpt_family="AT_rich"
repeat_region 18361. .18473
/rpt_family="(TAGA)n"
repeat_region complement(20017. .20206)
/rpt_family="MIR"
repeat_region complement(21062. .21181)
/rpt_family="L2"
repeat_region complement(23016. .23366)
/rpt_family="LIM4"
23392. .24560
/rpt_family="LIMEL"
repeat_region 24912. .24964
/rpt_family="AT_rich"
repeat_region 25816. .25929
/rpt_family="(CATG)n"
repeat_region 26195. .26342
/rpt_family="MIR"
repeat_region 26389. .26538
/rpt_family="L2"
repeat_region 26944. .27032
/rpt_family="L2"
27190. .27289
/rpt_family="L2"
repeat_region complement(27593. .27716)
/rpt_family="MIR"
repeat_region 29958. .30008
/rpt_family="(CA)n"
repeat_region complement(30009. .30162)
/rpt_family="MIR"
repeat_region 30231. .30515
/rpt_family="AluSg"
repeat_region complement(31648. .31999)
/rpt_family="MLT1A1"
32078. .32378
/rpt_family="AluSx"
repeat_region 32684. .32802
/rpt_family="MIR"
repeat_region 33334. .33381
```

Query Match 2.9%; Score 22; DB 9; Length 154850;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 22; Conservativity 0; Mismatches

0; Indels 0; Gaps 0;

QY 731 AATTAATAAATTTTATTTT 752

|||||

```
Db 17111 AATTAATAAATTTTATTTT 17090
RESULT 24
AL732407/c 158882 bp DNA linear HTG 18-MAY-2002
LOCUS Homo sapiens chromosome 10 clone RP11-50E23, *** SEQUENCING IN
DEFINITION PROGRESS ***; 5 unordered pieces.
ACCESSION AL732407
VERSION AL732407.4 GI:21065407
KEYWORDS HTG; PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 158882)
REFERENCE
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
on May 21, 2002 this sequence version replaced gi:20803707.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA50E23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 157683 bases at least Q40
Consensus quality: 158040 bases at least Q30
Consensus quality: 158293 bases at least Q20
Insert size: 158482; sum-of-contrigs
Insert size: 168042; 0.8% error; agarose-fp
Quality coverage: 7.51x in Q20 bases; sum-of-contrigs Quality
coverage: 7.13x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 40692: contrig of 40692 bp in length
* 1 40693: gap of 100 bp
* 40693 57562: contrig of 16770 bp in length
* 57563 57662: gap of 100 bp
* 57663 71113: contrig of 13451 bp in length
* 71114 71213: gap of 100 bp
* 71214 95237: contrig of 24024 bp in length
* 95238 95337: gap of 100 bp
* 95338 158882: contrig of 63345 bp in length.
FEATURES
source
1.158882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-50E23"
/clone.lib="RPC1-11.1"
1.40692
/note="assembly_fragment:01717
fragment_chain:1
clone_end:SP6
vector_side:left"
40793. .57562
/note="assembly_fragment:00699
fragment_chain:1"
57663. .71113
misc_feature
misc_feature
misc_feature
```

```
/note="assembly_fragment:02369
fragment_chain:1"
71214 .95237
misc_feature
/note="assembly_fragment:00597
fragment_chain:1"
95338 .158882
misc_feature
/note="assembly_fragment:01653
fragment_chain:1
clone_end:T7
vector_side:right"
BASE COUNT 43020 a 39921 c 37737 g 37803 t 401 others
ORIGIN
Query Match 2.9%; Score 22; DB 2; Length 158882;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 738 TAAATTTTATTTTATTTCT 759
|||||
Db 4758 TAAATTTTATTTTATTTCT 4737

RESULT 25
AC122884 160295 bp DNA linear ROD 16-JUL-2002
LOCUS Mus musculus clone RP23-88013, complete sequence.
DEFINITION AC122884
ACCESSION AC122884.2 GI:21844709
VERSION HTG.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 160295)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160295)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 160295)
REFERENCE McPherson,J.D. and Waterston,R.H.
AUTHORS Direct Submission
JOURNAL Submitted (16-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 16, 2002 this sequence version replaced gi:21307559.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M.BA0088013

FEATURES
Source Location/Qualifiers
1.160295
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-88013"

BASE COUNT 49461 a 27805 c 30083 g 52946 t
ORIGIN
Query Match 2.9%; Score 22; DB 10; Length 160295;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 AATAAATTAATTTTATTTT 752
|||||
Db 87979 AATAAATTAATTTTATTTT 87958
```

```
RESULT 26
AC069241 163838 bp DNA linear HTG 09-MAY-2002
LOCUS Homo sapiens chromosome 3 clone RP11-276K20, WORKING DRAFT
DEFINITION SEQUENCE, 12 unordered pieces.
ACCESSION AC069241
VERSION AC069241.8 GI:20335667
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 163838)
AUTHORS Muzny,D.M., Adams,C., Adio-Ogunola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,H.C., Ate,J.R., Ayale,M., Banks,T.,
Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Dejaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulvy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Koryan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabath,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwu,S., Ogih,M., Okunou,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Schreier,S., Scott,G., Shen,H., Shooshtari,N., Sison,I.,
Sodergren,E., Sonalkhe,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villano,D., Vinson,R., Wang,O.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
JOURNAL Unpublished
TITLE Direct Submission
JOURNAL 2 (bases 1 to 163838)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 163838)
REFERENCE Worley,K.C.
AUTHORS Direct Submission
JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Apr 28, 2002 this sequence version replaced gi:15887158.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
```



```
Assembly program: Phrap; version 0.990329
Consensus quality: 167300 bases at least Q40
Consensus quality: 169474 bases at least Q30
Consensus quality: 171145 bases at least Q20
Estimated insert size: 163866; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.bjsg.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      18529: contig of 18529 bp in length
*      18530: gap of unknown length
*      18630: 86341: contig of 67712 bp in length
*      86342: 86441: gap of unknown length
*      86442: 165120: contig of 78679 bp in length.
Location/Qualifiers
Source
1. .165120
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chrosome="3"
/clone="RP11-259F24"
BASE COUNT      52915 a 31909 c 31214 g 48881 t      201 others
ORIGIN
Query Match      2.9%; Score 22; DB 2; Length 165120;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 731 AATTAATAATTAATTTTATTTT 752
|||||
Db 67840 AATTAATAATTAATTTTATTTT 67861

RESULT 28
AL732395      167837 bp      DNA      linear      HTG 13-AUG-2002
LOCUS
DEFINITION
Homo sapiens chromosome 10 clone RP11-77F5, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
AL732395      AL732395.8      GI:22265446
VERSION
KEYWORDS
HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 167837)
Whitehead, S.
Direct Submissio
Submitted (10-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, G10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Aug 15, 2002 this sequence version replaced g1:21953064.
-----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
Project Information
Center project name: bA77F5
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; Big Dye; 18% of reads
Consensus quality: 166852 bases at least Q40
Consensus quality: 167090 bases at least Q30
Consensus quality: 167276 bases at least Q20
```

```
Insert size: 167837; sum-of-contigs
Insert size: 174475; 2.7% error; agarose-fp
Quality coverage: 13.52x in Q20 bases; sum-of-contigs Quality
coverage: 13.01x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
Source
1. .167837
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chrosome="10"
/clone="RP11-77F5"
/clone_lib="RPCT-11.1"
1. .167837
/note="assembly-fragment:02294
clone_end:SP6
vector_side:right"
misc-feature
BASE COUNT      39694 a 39959 c 42529 g 45655 t
ORIGIN
Query Match      2.9%; Score 22; DB 2; Length 167837;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 738 TAAATTTTATTTTATTTTCT 759
|||||
Db 158615 TAAATTTTATTTTATTTTCT 158636
```

```
RESULT 29
GTAJ10592      180915 bp      DNA      linear      PLN 26-APR-2001
LOCUS
DEFINITION
Guillardi theta DNA for complete sequence of nucleomorph
chromosome 2.
AJ010592
AJ010592.2      GI:12580654
18S ribosomal RNA; 18S rRNA gene; 26S proteasome; 28S ribosomal
RNA; 28S rRNA gene; 40S ribosomal protein S13; 40S ribosomal
protein S15A; 40S ribosomal protein S19; 40S ribosomal protein S23;
40S ribosomal protein S24; 40S ribosomal protein S26; 40S ribosomal
protein S27; 40S ribosomal protein S28; 40S ribosomal protein S3a;
40S ribosomal protein S4; 40S ribosomal protein S8; 40S ribosomal
protein SSA; 5.8S ribosomal RNA; 5.8S rRNA gene; 5S ribosomal RNA;
5S rRNA gene; 60S ribosomal protein L10; 60S ribosomal protein L13;
60S ribosomal protein L14; 60S ribosomal protein L15; 60S ribosomal
protein L21; 60S ribosomal protein L24; 60S ribosomal protein L27a;
60S ribosomal protein L30; 60S ribosomal protein L36; 60S ribosomal
protein L37A; 60S ribosomal protein L5; 60S ribosomal protein L8;
60S ribosomal protein L9; ATP-dependent RNA helicase; cap-binding
protein; cbdx gene; cbp gene; CCR4-associated factor; CDC2 gene;
cdc28 gene; cell division protein FtsZ; cpn60 gene; CPN60 protein;
DNA primase; DNA-directed RNA polymerase; dnag gene; elf1A gene;
ftsZ gene; gbp gene; gida gene; glucose inhibited division protein
A; gsp2 gene; GTP-binding nuclear protein RAN; GTP-binding protein;
H2B gene; has 1 gene; hct136 gene; histon H2b; lap100 gene; LAP100
enzyme; mRNA guanylyltransferase; N-myristoyl transferase; ncbbp2
protein; nml1 gene; nog1 gene; nop5 gene; nuclear cap binding protein;
nucleolar G-protein; nucleolar protein; p-bp gene; photosystem II;
pop2 gene; PrL1 gene; PRR8 gene; PRSA1 gene; PRSB1 gene; PRSB4
gene; PRSB1 gene; PRS10b gene; PRS12 gene; PRS13 gene; PRS54
gene; PRS6 gene; PRS7 gene; ranbp1 gene; rbp1 gene; Rev
interacting protein Rip-1; ribosomal protein CEP52; Rpl1 gene;
RNA-dependent helicase; rpl1 gene; rpl1 gene; RPB3 gene; rplc10
gene; rpl10 gene; rpl13 gene; rpl14 gene; rpl15 gene; rpl21 gene;
rpl14 gene; rpl27a gene; rpl30 gene; rpl16 gene; RPL37A gene; rpl140
gene; rpl5 gene; rpl8 gene; rpl9 gene; rps13 gene; rps24 gene;
rps17 gene; rps17 protein; rps19 gene; RPS23 gene; rps24 gene;
rps26 gene; rps27 gene; rps28 gene; RPS3a gene; RPS4 gene; rps8
gene; rtp3 gene; rtp4 gene; rub gene; rubredoxin; SAR DNA-binding
```

	protein-1; secE gene; sen1 gene; sen34 gene; serin-tRNA ligase; seeryl-tRNA synthetase; small nuclear ribonucleoprotein; small nuclear ribonucleoprotein E; small nuclear ribonucleoprotein SM D2; snRNP; snRPD2 gene; snRPD3 gene; snRPE gene; snul3 gene; sof1 gene; splicing factor Prpb; ste4+ gene; sys1 gene; T-complex protein; taf30 gene; taf9 gene; TARA binding protein; tcpg gene; tcpt gene; tcpg gene; tetrameric tRNA splicing endonuclease; tflia-s gene; tflIB related factor hbrf; tflIB-drif gene; tflID gene; tflI21 gene; transcription initiation factor IIA gamma chain; transcription initiation factor IID SU beta; transfer RNA-Arg; transfer RNA-Cys; transfer RNA-Gln; transfer RNA-His; transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer RNA-Phe; transfer RNA-Ser; transfer RNA-Thr; transfer RNA-Val; translation initiation factor; translational initiation factor 2 alpha SU; translocase; tRNA-Arg gene; tRNA-Cys gene; tRNA-Gln gene; tRNA-His gene; tRNA-Ile gene; tRNA-Leu gene; tRNA-Lys gene; tRNA-Phe gene; tRNA-Ser gene; tRNA-Thr gene; tRNA-Val gene; tuba gene; tubulin alpha-1 chain; U5 snRNP-specific protein; usnRNP gene; UBC4 gene; ubiquitin; ubiquitin conjugating enzyme; ubiquitin-conjugating enzyme E2-21 KD; uce-E2 gene.
SOURCE	Guillardia theta.
ORGANISM	Guillardia theta.
REFERENCE	Eukaryota; Cryptomonadaceae; Guillardia.
AUTHORS	1 (bases 1 to 180915) Zauner,S., Fraunholz,M., Wastl,J., Penny,S., Beaton,M., Cavalier-Smith,T., Maier,U.G. and Douglas,S.
TITLE	Chloroplast protein and centriosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (1), 200-205 (2000)
MEDLINE	20087226
PUBMED	10618395
REFERENCE	2 (bases 1 to 180915) Douglas,S., Zauner,S., Fraunholz,M., Beaton,M., Penny,S., Deng,L.T., Wu,X., Reith,M., Cavalier-Smith,T. and Maier,U.G.
AUTHORS	The highly reduced genome of an enslaved algal nucleus
JOURNAL	Nature 410 (6832), 1091-1096 (2001)
MEDLINE	21223349
PUBMED	11323671
REFERENCE	3 (bases 1 to 180915) Zauner,S.
AUTHORS	Direct Submission
TITLE	Submitted (27-AUG-1998)
JOURNAL	Cell Biology and Applied Botany, Karl-von-Frisch-Stasse, D-35032 Marburg, GERMANY
REMARK	Revised by [4]
REFERENCE	4 (bases 1 to 180915) Zauner,S.
AUTHORS	Direct Submission
TITLE	Submitted (23-JAN-2001)
JOURNAL	Zauner S., Philipps-University Marburg, Cell Biology and Applied Botany, Karl-von-Frisch-Stasse, D-35032 Marburg, GERMANY
COMMENT	On Jan 26, 2001 this sequence version replaced gi:4583655.
FEATURES	Location/Qualifiers
source	1..180915 /organism="Guillardia theta" /db_xref="taxon:55529" /chromosome="nucleomorph chromosome 2"
gene	279..398
rRNA	/gene="5S rRNA" 279..398
gene	/gene="5S rRNA" /product="5S ribosomal RNA" complement(936..4934)
rRNA	/gene="28S rRNA" complement(936..4934)
gene	/gene="28S rRNA" /product="28S ribosomal RNA" complement(5493..5663)
rRNA	/gene="5.8S rRNA" complement(5493..5663)
gene	/gene="5.8S rRNA" /product="5.8S ribosomal RNA"
gene	complement(6175..8214) /gene="18S rRNA" complement(6175..8214) /gene="18S rRNA" /product="18S ribosomal RNA" 9177..9509
rRNA	/note="ORF110" /codon_start=1 /product="hypothetical protein" /protein_id="CAC26972.1" /db_xref="GI:12580655"
CDS	/translation="MKNTKGKTRAGKPKNDLTTRTPRAPIDEPNRGKCTAHLSLPPMSKRSSAKMILPGQGVLNKGTSRDCKPDLPMATSSCAPISTERLPMRMAWDTRELNS" 9701..10333 /note="ORF210" /codon_start=1 /product="hypothetical protein" /protein_id="CAC26973.1" /db_xref="GI:12580656"
CDS	/translation="MVRIFKRRMGMMVQPTSRGSGTHLRLILKTRHGLNNYRIQHPRGSPCAERTVSNRCBERTPLRTKIVTTLQGITSTPREMWSQSLSPSSSYE EARGEDMMPRRPSCSPUNMLRPEESYSTHTDSDNTRANIRBGHOKSSDGKRS RPKTISESHNIQOKRTIRLRLHTSPLORLKGGTSSRHTLPLTKPLGV" complement(10314..10769) /note="ORF151" /codon_start=1 /product="hypothetical protein" /protein_id="CAC26974.1" /db_xref="GI:12580657"
CDS	/translation="MRRRCVSYPPRYFFFLFFFPFWRESSLVFLFFFLAHEEVLVCDFVFWRLFPPLGSLILINVLPLGCGOGRGSSWVVERCVISFFLILSLRGRMSLDRMKCKRKACPSLRGERRSSSLSPRWVGCISLHLSLRHPV" 10833..11030 /note="ORF65" /codon_start=1 /product="hypothetical protein" /protein_id="CAC26975.1" /db_xref="GI:12580658"
CDS	/translation="MTDRPSCHPSPVFNFYQKKEMFTHSACKOYTRMPDSNERSKTHLSHPESIDQFTITKQN" 11187..11396 /note="ORF69" /codon_start=1 /product="hypothetical protein" /protein_id="CAC26976.1" /db_xref="GI:12580659"
CDS	/translation="MIPAARKIIPLEKRPPOPLPAPFTNPGKKERMSRFPPOGRNRTGANTRTRPSOKIKARSTPGSL" 12264..12707 /gene="ubc4" 12264..12707 /gene="ubc4" /codon_start=1 /product="ubiquitin conjugating enzyme" /protein_id="CAC26977.1" /db_xref="GI:12580660"
gene	/db_xref="GI:12580660" /translation="MATRKIOKEHKDLKDPNSCAGSPDENLFRWTASTIGSDSPYAGGVFFSLTFPPDYPPKPKVOFTTKTIYIPNINSNGSICLIDLKDOMSPALTVSKY LLSLISLTLDPNPDPPLVPELTAIVFKENRSRYEATARWORYAT" 13171..13920 /gene="tflid" 13171..13920 /gene="tflid" /codon_start=1 /product="TATA binding protein of transcription factor IID" /protein_id="CAB40395.1" /db_xref="GI:4583657"
CDS	/db_xref="GI:4583657" /translation="WICKSEKNSIQGISQSDIDILIRKTNNTSNVFFQGNLHAF DVIRRGSTEAQSSVYPERAVRANETIPNIVSTVSLGIDLDLALAKARNAEY NPRFAAVIMRIRDPKPTALIFSSGKMWVTAKSSEDSARVACKRYARIQRLGYHAK

CDS

FIDRIONIVASCDVPRPRLRSLAHNAHOCSEPELPGLIYRMITPKVILLIEVS
GKLVTCAGKORNDIPOAFSNISYSLICYKKT"
complement(114068..16089)
/note="ORF673"
/codon_start=1
/product="hypothetical protein"
/protein_id="CA840396.1"
/db_xref="GI:4583658"
/db_xref="SPTRMBL:O9XG31"
/translation="MKITVLYFENNFKELFDLYFINPNKLIDKSLFLIKINTWKY
DPLIVKCLCKPIFEISLSNIIITNEFTNEIKIMFEKLKIKFGSLKSVNLE
KLKFEINVTWLCFLREIQRLKSNYIVLYDONRKNITYSSEETKYRSL
LKNSSLRIEYGLSIEFPFNIIISANDLSKKNHILADSDFEESCIIHYKRVGILCP
LNSLKQFIKNFPEFLKASIVITTKNNYSRNSRPSFPEFKLAKSPKSLSSFPY
PYNLSRLIKKIIILSSIGTKRISYDINSIQRLKLFPCASIEYKQILRIYTNLT

Query Match 2.9%; Score 22; DB 8; Length 180915;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 736 AATAAAATTTTATTATTATT 757
Db 115190 AATAAAATTTTATTATTATT 115211
|||||

RESULT 30
AL365207/ 182679 bp DNA linear PRI 01-MAR-2001
LOCUS Human DNA sequence from clone RP11-420024 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL365207
VERSION AL365207.17 GI:13234948
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 182679)
Blakey, S.
Direct Submission
Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonequery@sanger.ac.uk
On Mar 5, 2001 this sequence version replaced gi:13184362.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chri
RP11-420024 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-420024 The true
left end of clone RP11-417K10 is at 104893 in this sequence. The
true right end of clone RP11-336D15 is at 51737 in this sequence.

FEATURES
source

Location/Qualifiers
1..182679
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-420024"
/clone_1fb="RPCI-11.2"
1..1120
/note="HERVL repeat: matches 4537..5654 of consensus"
1121..1247
/note="LTR2A repeat: matches 1..129 of consensus"
1248..1592
/note="THER1B repeat: matches 1..364 of consensus"
1593..1755
/note="LTR2A repeat: matches 129..286 of consensus"
1756..2465
/note="L1P repeat: matches 3899..4611 of consensus"
2672..2705
/note="L17 copies 2 mer tg 97% conserved"
3180..3795
/note="L1MBC repeat: matches 290..891 of consensus"
3802..4112
/note="L1ME repeat: matches 5185..5508 of consensus"
5219..5326
/note="L1MD1 repeat: matches 6108..6217 of consensus"
5563..5790
/note="L1MA8 repeat: matches 5662..5793 of consensus"
5826..6332
/note="L1MA8 repeat: matches 5783..6289 of consensus"
8213..8325
/note="L2 repeat: matches 2593..2710 of consensus"
8723..8898
/note="88 copies 2 mer tt 56% conserved"
9354..9544
/note="L1MD2 repeat: matches 5678..5883 of consensus"
9560..9773
/note="MER20 repeat: matches 1..198 of consensus"
9860..10412
/note="L1MD2 repeat: matches 4984..5557 of consensus"
10848..11340
/note="L2 repeat: matches 1855..2352 of consensus"
complement(11343..11848)
/note="match: GSS: Em:AQ769355"
complement(11710..12057)
/note="match: GSS: Em:AQ037488"
11866..12232
/note="match: GSS: Em:AQ452860"
11897..12442
/note="match: GSS: Em:AQ285658"
12082..12457
/note="match: GSS: Em:AQ119499"
14318..14461
/note="LTR1H repeat: matches 390..528 of consensus"
16598..16939
/note="L1M5 repeat: matches 1169..1423 of consensus"
17463..17661
/note="MER20 repeat: matches 1..213 of consensus"
18050..18077
/note="L14 copies 2 mer tt 89% conserved"
18078..18267
/note="LTR1C repeat: matches 221..420 of consensus"
18303..18951
/note="L1PA5 repeat: matches 5498..6143 of consensus"
19487..19810
/note="AluY repeat: matches 1..309 of consensus"
21024..21131
/note="L2 repeat: matches 2302..2408 of consensus"
21132..21442
/note="AluY repeat: matches 1..311 of consensus"
21443..22006
/note="L2 repeat: matches 1732..2302 of consensus"

```

repeat_region 22153..22190
/note="19 copies 2 mer tg 78% conserved"
repeat_region 22328..22612
/note="Alub repeat: matches 1..281 of consensus"
repeat_region 23948..24144
/note="MER58A repeat: matches 21..224 of consensus"
repeat_region 25151..25287
/note="LTR33 repeat: matches 138..286 of consensus"
repeat_region 25419..25940
/note="LTR1D repeat: matches 1..505 of consensus"
repeat_region 26711..27083
/note="MER67C repeat: matches 330..710 of consensus"
misc_feature 26873..27367
/note="match: GSS: Em:AQ777120"
repeat_region 27374..27767
/note="LTR1 repeat: matches 4..410 of consensus"
misc_feature complement(27543..27936)
/note="match: GSS: Em:AQ086254"
repeat_region 29359..31285
/note="LIMB2 repeat: matches 4236..6160 of consensus"
repeat_region 31963..31992
/note="L15 copies 2 mer aa 86% conserved"
repeat_region 34044..34097
/note="27 copies 2 mer tg 77% conserved"
repeat_region 34375..34412
/note="MIR repeat: matches 67..105 of consensus"
repeat_region 35211..35386
/note="Aluo repeat: matches 113..288 of consensus"
repeat_region 36001..37974
/note="LIP4 repeat: matches 4169..6142 of consensus"
repeat_region 37974..39103
/note="LIP4 repeat: matches 3044..4175 of consensus"
misc_feature complement(39766..40090)
/note="match: STS: Em:G09144"
repeat_region 39916..39965
/note="25 copies 2 mer ga 76% conserved"
misc_feature complement(40658..41036)
/note="match: GSS: Em:AQ069119"
repeat_region 42121..42222
/note="Tiger3b repeat: matches 6..102 of consensus"
repeat_region 42214..43101
/note="Tiger3b repeat: matches 266..1230 of consensus"
repeat_region 45786..46146
/note="LIM4 repeat: matches 5910..6277 of consensus"
repeat_region 46904..47282
/note="L2 repeat: matches 2345..2750 of consensus"
repeat_region 47328..48022
/note="MER50 repeat: matches 3..711 of consensus"
repeat_region 48024..48300
/note="L2 repeat: matches 1930..2264 of consensus"
repeat_region 48355..48519
/note="LTR41 repeat: matches 1..159 of consensus"
repeat_region 49185..49777
/note="MER49 repeat: matches 214..825 of consensus"
repeat_region 50041..50488
/note="L2 repeat: matches 395..848 of consensus"
repeat_region 51275..51548
/note="L1 repeat: matches 3461..3724 of consensus"
repeat_region 51640..52198
/note="LIM10 repeat: matches 5182..5822 of consensus"
repeat_region 52206..52499
/note="Aluc repeat: matches 1..294 of consensus"
repeat_region 52501..52544
/note="22 copies 2 mer tc 79% conserved"
repeat_region 52546..53065
/note="SVA repeat: matches 881..1386 of consensus"
misc_feature 52735..53555
/note="Cpg island"
evidence-not_experimental
repeat_region 53067..53337
/note="SVA repeat: matches 725..993 of consensus"
repeat_region 53339..54274
/note="SVA repeat: matches 9..955 of consensus"

```

```

repeat_region 54276..54440
/note="LTR1A2 repeat: matches 161..324 of consensus"
repeat_region 54472..54908
/note="LIM10 repeat: matches 5871..6319 of consensus"
repeat_region 54990..55066
/note="MIR repeat: matches 168..245 of consensus"
misc_feature 55396..56091
/note="match: GSS: Em:AQ119091"

Query Match 2.9% Score 22; DB 9; Length 182679;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 AATTAATAAATTTTATTTT 752
Db 51956 AATTAATAAATTTTATTTT 51935

RESULT 31
AC122485/c
LOCUS
DEFINITION
Mus musculus chromosome UNK clone RP24-377J10, WORKING DRAFT
ACCESSION
AC122485.1 GI:21105949
VERSION
AC122485.1
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 183594)
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
The sequence of Mus musculus clone
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 183594)
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0372J10
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 181199 bases at least Q40
Consensus quality: 181717 bases at least Q30
Consensus quality: 182035 bases at least Q20
Insert size: 165000; agarose-1p
Insert size: 182556; sum-of-ctligs
Quality coverage: 14.02 in Q20 bases; agarose-1p
Quality coverage: 11.39 in Q20 bases; sum-of-ctligs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 538: contig of 538 bp in length
* 539 638: gap of unknown length
* 639 2716: contig of 2078 bp in length

```

	FEATURES	source
*	2717	2816: gap of unknown length
*	2817	4835: contig of 2019 bp in length
*	4836	4935: gap of unknown length
*	4936	157922: contig of 10857 bp in length
*	15793	15892: gap of unknown length
*	15693	59894: contig of 44002 bp in length
*	59893	59994: gap of unknown length
*	59995	183594: contig of 123600 bp in length.
	Location/Qualifiers	
	1..183594	
	/organism="Mus musculus"	
	/db_xref="taxon:10090"	
	/chromosome="UNK"	
	/clone="RP24-377j10"	
	1..538	
misc_feature	/note="assembly_name:Contig15"	
	639..2716	
misc_feature	/note="assembly_name:Contig37"	
	2817..4835	
misc_feature	/note="assembly_name:Contig38"	
	4936..15792	
misc_feature	/note="assembly_name:Contig39"	
	15893..59894	
misc_feature	/note="assembly_name:Contig40"	
	59995..183594	
misc_feature	/note="assembly_name:Contig41"	
	59458 a 35424 c 35554 g 52657 t	501 others

Query Match	2.9%	Score 22	DB 2	Length 183594
Best Local Similarity	100.0%	Pred. No. 3.6		
Matches 22	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy 738 TAAATTTTATTATTCT 759
 |||||
 Db 131622 TAAATTTTATTATTCT 131601

RESULT 32	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AC103310	249289 bp	linear	AC103310	AC103310.4	HTG, HTGS_PHASE1.	Norway rat.	Rattus norvegicus
		*** SEQUENCING IN PROGRESS					
		***, 59 unordered pieces.					

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
AUTHORS

1 (bases 1 to 249289)

Muzny, D.M., Adams, C., Adio-Oduola, B., Alt-osman, F.R., Allen, C., Alstirooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbaria, J., Benton, J., Blinage, K., Blankenbump, K., Bonan, D., Bouck, J., Bowe, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Cahay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Burren, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dedertlich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.J., Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotco, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homi, F., Howard, S., Huber, J., Hulys, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratavicz, J., Kurshil, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.

TITLE	Direct Submission	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 249289)	
AUTHORS	Worley,K.C.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-NOV-2001)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 249289)	
AUTHORS	Worley,K.C.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-JUL-2002)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

Li J., Li L., Licharge O., Lieu C., Liu J., Liu M., Loutsched, H.,
Lozano R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Madsen H., Mader, M., Mapia P., Martin R., Martindale A., Martinez E.,
Massy E., Maxhimer E., McLeod M.P., Meador M., Mei G., Metzger M.,
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
Moser M., Neal D., Newton J., Newton S., Ogih M., Okunishi N.,
Nguyen N., Nickerson E., Nwokwenkwo S., Ogun M., Okunishi N.,
Oragunye N., Oyler R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Pitman E., Pu L.L., Quiles M., Ren Y.,
Rivers M., Rojas A., Rojibokun I., Rolfe M., Ruiz S., Savary G.,
Scheer S., Scott G., Shen H., Shooshbari N., Sisson I.,
Sodergren E., Sonalke T., Sparks A., Stanley H., Stone H.,
Sutton A., Svatek A., Taber P., Tameris A., Tameris K., Tang H.,
Tansay J., Taylor C., Taylor T., Teiford B., Thomas N., Thomas S.,
Usmani K., Vasquez L., Vera V., Villalón D., Vinson R., Wang Q.,
Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
Williams G., Williamson A., Wleciyk R., Wooden S., Worley K.,
Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,
Weinstock G. and Gibbs R.

Direct Submission
Unpublished
2 (bases 1 to 249289)
Worley, R.C.

Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 249289)
Worley, R.C.

Direct Submission
Submitted (14-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi.18700889.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GKES
Center clone name: CH230-193012
----- Summary Statistics -----
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 190767 bases at least Q40
Consensus quality: 194243 bases at least Q30
Consensus quality: 196500 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1430: contig of 1430 bp in length
1431 1530: gap of unknown length
1531 2669: contig of 1139 bp in length
2670 2769: gap of unknown length
2770 3998: contig of 1230 bp in length
4000 4099: gap of unknown length
5244: contig of 1145 bp in length
5245 5344: gap of unknown length
5345 6448: contig of 1104 bp in length
6449 6548: gap of unknown length
6549 7574: contig of 1026 bp in length
7575 7674: gap of unknown length
7675 8903: contig of 1229 bp in length

```
*      8904      9003: gap of unknown length
*      9004      10020: contig of 1017 bp in length
*     10021      10120: gap of unknown length
*     10121      11403: contig of 1283 bp in length
*     11404      11503: gap of unknown length
*     11504      12831: contig of 1328 bp in length
*     12832      12931: gap of unknown length
*     12932      14049: contig of 1118 bp in length
*     14050      14149: gap of unknown length
*     14150      15301: contig of 1152 bp in length
*     15302      15401: gap of unknown length
*     15402      16597: contig of 1196 bp in length
*     16598      16697: gap of unknown length
*     17715      17714: contig of 1017 bp in length
*     17815      19053: contig of 1239 bp in length
*     19054      19153: gap of unknown length
*     19154      20657: contig of 1504 bp in length
*     20658      20757: gap of unknown length
*     20758      21786: contig of 1029 bp in length
*     21787      21886: gap of unknown length
*     21887      23057: contig of 1171 bp in length
*     23058      23157: gap of unknown length
*     23158      24412: contig of 1255 bp in length
*     24413      24512: gap of unknown length
*     24513      25786: contig of 1274 bp in length
*     25787      25886: gap of unknown length
*     25887      26954: contig of 1068 bp in length
*     26955      27054: gap of unknown length
*     27055      28058: contig of 1004 bp in length
*     28059      28158: gap of unknown length
*     28159      29213: contig of 1055 bp in length
*     29214      29313: gap of unknown length
*     29314      30454: contig of 1141 bp in length
*     30455      30554: gap of unknown length
*     30555      31367: contig of 1413 bp in length
*     31368      32067: gap of unknown length
*     32068      33153: contig of 1086 bp in length
*     33154      33253: gap of unknown length
*     33254      34527: contig of 1274 bp in length
*     34528      34627: gap of unknown length
*     34628      35778: contig of 1151 bp in length
*     35779      35878: gap of unknown length
*     35879      37154: contig of 1276 bp in length
*     37155      37254: gap of unknown length
*     37255      38433: contig of 1179 bp in length
*     38434      38533: gap of unknown length
*     38534      40118: contig of 1585 bp in length
*     40119      40218: gap of unknown length
*     40219      41245: contig of 1027 bp in length
*     41246      41345: gap of unknown length
*     41346      43777: contig of 2432 bp in length
*     43778      43877: gap of unknown length
*     43878      46124: contig of 2247 bp in length
*     46125      46224: gap of unknown length
*     46225      47724: contig of 1500 bp in length
*     47725      47824: gap of unknown length
*     47825      49083: contig of 1259 bp in length
*     49084      49183: gap of unknown length
*     49184      50409: contig of 1226 bp in length
*     50410      50509: gap of unknown length
*     50510      53021: contig of 2512 bp in length
*     53022      53121: gap of unknown length
*     53122      55591: contig of 2470 bp in length
*     55592      55691: gap of unknown length
*     55692      59070: contig of 3379 bp in length
*     59071      59170: gap of unknown length
*     59171      63524: contig of 4354 bp in length
*     63525      63624: gap of unknown length
*     63626      69326: contig of 5602 bp in length
*     69327      69326: gap of unknown length
*     69327      74137: contig of 4811 bp in length
*     74138      74237: gap of unknown length
```

```
*      74238      79949: contig of 5712 bp in length
*      79950      80049: gap of unknown length
*      80050      85156: contig of 5107 bp in length
*      85157      85256: gap of unknown length
*      85257      92434: contig of 7178 bp in length
*      92435      92534: gap of unknown length
*      92535      100244: contig of 7710 bp in length
*      100245      100245: gap of unknown length
*      100345      106991: contig of 6647 bp in length
*      106992      107091: gap of unknown length
*      107092      113892: contig of 6801 bp in length
*      113893      113992: gap of unknown length
*      113993      122519: contig of 8527 bp in length
*      122520      122620: gap of unknown length
*      131392      131392: contig of 8773 bp in length
*      131393      131492: gap of unknown length
*      131493      143228: contig of 11736 bp in length
*      143229      143229: gap of unknown length
*      143329      156394: contig of 13066 bp in length
*      156395      156494: gap of unknown length
*      156495      168854: contig of 12360 bp in length

Query Match      2.9%; Score 22; DB 2; Length 249289;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      736 AATAAATTTTATTTTATTTT 757
Db 25348 AATAAATTTTATTTTATTTT 25369

RESULT 33
AX028709/c      28 bp      DNA      linear      PAT 16-SEP-2000
DEFINITION      Sequence 13 from Patent EP1018550.
ACCESSION      AX028709
VERSION      AX028709.1 GI:10189822
KEYWORDS
SOURCE
ORGANISM      .
European house dust mite.
Dermatophagoides pteronyssinus
Eukaryotes; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcoptiformes; Astigmata; Pyroglyphidae;
Dermatophagoides.
REFERENCE      1 (bases 1 to 28)
AUTHORS      Thomas,W.R. and Chua,K.Y.
TITLE      Allergenic protein and peptides from house dust mite and uses
thereof
JOURNAL      Patent: EP 1018550-A 13 12-JUL-2000;
INST CHILD HEALTH RESEARCH (AU)
FEATURES
source      1..28
Location/Qualifiers
BASE COUNT      7 a      5 c      3 g      13 t
ORIGIN
Query Match      2.8%; Score 21; DB 6; Length 28;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      664 CGTGAATTGAAAAAATTTAA 684
Db      28 CGTGAATTGAAAAAATTTAA 8

RESULT 34
AF289137/c      493 bp      DNA      linear      INV 29-APR-2002
LOCUS      Aphelinus asychis 16S ribosomal RNA gene, partial sequence;
DEFINITION      mitochondrial gene for mitochondrial product.
ACCESSION      AF289137
VERSION      AF289137.1 GI:15071127
KEYWORDS
```

SOURCE Aphelinus asychis.
ORGANISM Mitochondrion Aphelinus asychis
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apocrita; Chalcidoidea;
Aphelinidae; Aphelinus.
REFERENCE 1 (bases 1 to 493)
AUTHORS Chen, Y., Giles, K.L. and Greenstone, M.H.
TITLE Molecular Evidence for a Species Complex in the Genus Aphelinus
(Hymenoptera: Aphelinidae), with Additional Data on Aphidline
Phylogeny (Hymenoptera: Braconidae)
JOURNAL Ann. Entomol. Soc. Am. 95 (1), 29-34 (2002)
REFERENCE 2 (bases 1 to 493)
AUTHORS Chen, Y., Giles, K.L. and Greenstone, M.H.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2000) Department of Entomology and Plant
Pathology, Oklahoma State University, 127/110, Noble Research
Center, Stillwater, OK 74075, USA
FEATURES
source Location/Qualifiers
1..493
/organism="Aphelinus asychis"
/organelle="mitochondrion"
/db_xref="taxon:163153"
/country="Kazakhstan"
rRNA
/product="16S ribosomal RNA"
BASE COUNT 196 a 33 c 50 g 214 t
ORIGIN
Query Match 2.8%; Score 21; DB 3; Length 493;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 730 CAATAAATAAATTTTATT 750
|||||
Db 284 CAATAAATAAATTTTATT 264
RESULT 35
AF289138 493 bp DNA linear INV 29-APR-2002
LOCUS Aphelinus asychis 16S ribosomal RNA gene, partial sequence;
DEFINITION mitochondrial gene for mitochondrial product.
ACCESSION AF289138
VERSION AF289138.1 GI:15077128
KEYWORDS
SOURCE Aphelinus asychis.
ORGANISM Mitochondrion Aphelinus asychis
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apocrita; Chalcidoidea;
Aphelinidae; Aphelinus.
REFERENCE 1 (bases 1 to 493)
AUTHORS Chen, Y., Giles, K.L. and Greenstone, M.H.
TITLE Molecular Evidence for a Species Complex in the Genus Aphelinus
(Hymenoptera: Aphelinidae), with Additional Data on Aphidline
Phylogeny (Hymenoptera: Braconidae)
JOURNAL Ann. Entomol. Soc. Am. 95 (1), 29-34 (2002)
REFERENCE 2 (bases 1 to 493)
AUTHORS Chen, Y., Giles, K.L. and Greenstone, M.H.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2000) Department of Entomology and Plant
Pathology, Oklahoma State University, 127/110, Noble Research
Center, Stillwater, OK 74075, USA
FEATURES
source Location/Qualifiers
1..493
/organism="Aphelinus asychis"
/organelle="mitochondrion"
/db_xref="taxon:163153"
/country="France: Montpellier"
rRNA
/product="16S ribosomal RNA"
BASE COUNT 196 a 34 c 50 g 213 t
ORIGIN

Query Match 2.8%; Score 21; DB 3; Length 493;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 730 CAATAAATAAATTTTATT 750
|||||
Db 284 CAATAAATAAATTTTATT 264
RESULT 36
AB032753 509 bp mRNA linear PLN 12-JAN-2000
LOCUS Solanum melongena EEF13 mRNA, complete cds.
DEFINITION Solanum melongena EEF13 mRNA, complete cds.
ACCESSION AB032753
VERSION AB032753.1 GI:6691512
KEYWORDS
SOURCE Solanum melongena (strain:Senryo No. 2) pollinated fruit at 3 days
after anthesis ovary/fruit cDNA to mRNA.
ORGANISM Solanum melongena
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; easterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 509)
AUTHORS Nagasawa, M., Mori, H. and Yamaki, S.
TITLE Genes preferentially expressed at early stage of fruit set in
eggplant
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 509)
AUTHORS Nagasawa, M., Mori, H. and Yamaki, S.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) Masaki Nagasawa, Nagoya University,
Graduate School of Bioagricultural Sciences, Furocho Chikusa-ku,
Nagoya, Aichi 464-8601, Japan
(E-mail:1992013@embox.media.nagoya-u.ac.jp, Tel:+81-52-789-4028,
Fax:+81-52-789-4025)
FEATURES
source Location/Qualifiers
1..509
/organism="Solanum melongena"
/strain="Senryo No. 2"
/db_xref="taxon:4111"
/tissue_type="ovary/fruit"
/dev_stage="pollinated fruit at 3 days after anthesis"
gene 1..509
/gene="EEF13"
159..383
/gene="EEF13"
/codon_start=1
/protein_id="BA89332.1"
/db_xref="GI:6691513"
/translation="MATIAKKRRNKGIMLYKRIDSGSANQFSSKRTDLYKAGRYIV
GNIGDDWTDLIGENVGAKTFKVPDPMTIG"
BASE COUNT 183 a 86 c 109 g 131 t
ORIGIN
Query Match 2.8%; Score 21; DB 8; Length 509;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 731 AATAAATAAATTTTATT 751
|||||
Db 443 AATAAATAAATTTTATT 463
RESULT 37
GB1302036 2171 bp mRNA linear INV 12-DEC-2001
LOCUS Gryllus bimaculatus partial mRNA for allatostatin A prohormone
DEFINITION Gryllus bimaculatus partial mRNA for allatostatin A prohormone
precursor (ast gene).
ACCESSION AJ302036
VERSION AJ302036.1 GI:17736686
KEYWORDS allatostatin A prohormone precursor; ast gene.
SOURCE two-spotted cricket.
ORGANISM Gryllus bimaculatus

REFERENCE 1 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Ensifera; Gryllidae; Gryllinae; Gryllus.

AUTHORS 1 Meyerling-Vos,M., Wu,X., Huang,J., Jindra,M., Hoffmann,K.H. and Sehmal,F.

TITLE The allatostatin A gene of the cricket Gryllus bimaculatus (Ensifera, Gryllidae)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2171)

AUTHORS Meyerling-Vos,M.

TITLE Direct Submission

JOURNAL Submitted (14-DEC-2000) Meyerling-Vos M., Animal ecology I, University of Bayreuth, Universitaetsstr. 30, 95440 Bayreuth, GERMANY

FEATURES

source Location/Qualifiers

1..2171

organism="Gryllus bimaculatus"

db_xref="taxon:6999"

1..1038

gene="ast"

<1..1038

gene="ast"

function="release neuropeptides"

codon_start=1

product="allatostatin A prohormone precursor"

protein_id="CAC83078.1"

db_xref="GI:17736687"

translation="PASPDAAOEAGAELEPTEMEAGSGATPDELEFERYKLTDPGV GKRASTYSEKRLFYINFGGLKRRGKRGYGLKRRGROYGGLGRTGDEDDY YFPDEEDVPEEDNLDDSDSYDKRDRLVSFGKRSRPGFGLGRAGYSGLKERA OHQYSGFGLKRGEGMSYFGLKRPNYEMASRPNGFGLGRADANPYLLSDLGEEK RGPDRFAFGLKREVSNLEAVREEDLHDKEAQOHELAAPAPEREPEPDAAANG KHAKRSLHYGIGIKRTSDARGLDVPEEDDRDAISDFYIRPYSFGLGKRVPM YDFGIGRADR"

transit_peptide <1..108

gene="ast"

115..132

mat_peptide /gene="ast"

product="allatostatin 1"

142..195

mat_peptide /gene="ast"

product="allatostatin 2"

205..228

mat_peptide /gene="ast"

product="allatostatin 3"

244..273

mat_peptide /gene="ast"

product="allatostatin 4"

379..402

mat_peptide /evidence="experimental"

412..435

mat_peptide /gene="ast"

product="allatostatin 5"

445..468

mat_peptide /evidence="experimental"

478..504

mat_peptide /evidence="experimental"

514..543

mat_peptide /evidence="experimental"

553..597

mat_peptide /product="allatostatin 9"

847..870

mat_peptide /product="allatostatin 10"

Query Match 2.8%; Score 21; DB 3; length 2171; Best Local Similarity 100.0%; Pred. No. 18; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 733 TAAATPAAATTTTATTTT 753

Db 2130 TAAATPAAATTTTATTTT 2110

RESULT 38

LOCUS AF07803885

DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial cds, and cele gene, complete cds.

ACCESSION AF078042

VERSION AF078042.1 GI:12743878

KEYWORDS

SEGMENT

SOURCE

ORGANISM

1. 5 of 5

Caldicellulosiruptor sp. Tok7B.1.

Caldicellulosiruptor sp. Tok7B.1.

Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.

1 (bases 1 to 6005)

Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P., Williams,D.P. and Bergquist,P.L.

Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1

Curr. Microbiol. 40 (5), 333-340 (2000)

20171169

JOURNAL MEDLINE 10706665

PUBMED

REFERENCE 2 (bases 1 to 6005)

AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P., Williams,D.P. and Bergquist,P.L.

Direct Submission

Submitted (27-OCT-1999) Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

TITLE Location/Qualifiers

JOURNAL

source

1..6005

organism="Caldicellulosiruptor sp. Tok7B.1"

isolate="Tok7B.1"

db_xref="taxon:80339"

<1..506

note="ORF6; multidomain"

codon_start=3

transl_table=11

product="glycosyl hydrolase 6"

protein_id="AAK06393.1"

db_xref="GI:12743884"

translation="GQPDYNGTYGNPNLHKIVDYGNDIGITFASLANALLYSAAAT KEVYSDAARAKLAKELIDRMWNLYRDDKGISAPKRGDYRFFQEVYIPAGWTGKM PNGDIKSGVKRTIDIRSKYKQDDPMQKLSAYNAGEAFEPRIHRMACDIAIANATY EILFNGN"

<1..503

note="Region: family 48 glycosyl hydrolase domain"

617..5872

note="multidomain glycosyl hydrolase"

codon_start=1

evidence="experimental"

misc_feature

CDS

```
/transl_table=11
/product="Celle"
/protein_id="AAK06394.1"
/db_xref="GI:12743885"
/translation="MOEMKARKRVYSITALLVLTSLCFPGIMPKVAKAGTYWYKGA
LOKTIWYEROMSGKLPSSVNNRNGSGLDGQDGLDITLGGHMDGDKHFLPKMS
YSASMLGMAWYVEYKDAFVSKQLEHLLNQLEMAWYKSPKRSYVYVYOGDPTVH
NFMGPAVOMKRPAYKCDLSPASSVVAEASIAVASVYIKRNSOKASVYLOHK
DLFEFADTTRSDGATATGTYTSGGFLDGLMAAVMLYIATNDSSYLTKAEELMEY
ANGNTWTCOCDDVRYGTLMLAKITGKELKGAVERNLDMWDRITTYTPKMAVILG
WGSIRYATTAFLACVYADNSGCDNSNKTIKLNPAAKSQIDALSTGRSPVYGCTIN
PQPHHRNASHSSWANSMSKIPYHRIILYGLAVGGSDSDINDITTYVQVEVACDN
AGIYALAKAMQILGGEPIDEFKALIEPTNDELPESEKFGSGSPNTEVYSYNT
GWPRVYDKLSFKFYIDLTELQAGYSPDVYKVDYTYIEGKISGYPWDMNRLYY
LVDSGFKIRYDGEVEKHKQAFKISVPOGPMPTNDPNSKGLSOLERKKYIAAD
NNNLWMLGEPGAATSPAPSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
YKNNETASSTGSIIPRKIVNGSSSYDLSVYKIRWYTDGDKPQSAVCDMAIGAS
NVTPEVSKLSGVSADYILEVGFSSGAGQLQPKDPTGDQVREKNDMSVYNOADM
SMLQSMNNGENAKVTLVYDGLVWQGPAGGATPAPTATPTPTPTPTPTPTPTPT
TPVYATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
TAVDSLHVRYVYITIDGATOSVSVASINPAYIDVKEVGLAGAGADYVIEGK
SGAGVLAAGOSTKREIRLSIOKSGSVNOSNDYSVASNYIENKRYGIDYVIMKR
EPGRNADIKWYANGNLSMTNVLNPKIKLENGTAVDSRVKRVRYTYIDGATVS
VSVYSSINPAYIDVKEVGLAGAGADYVIEGKSGAGVLAAGOSTKREIRLSIOKSG
SVYSSNDYSVASNYIENKRYGIDYVIMKRSGAGVLAAGOSTKREIRLSIOKSG
PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
PARLIGNRLGTYMENNMSAGSDWHSDDVYCYIMGTGNDKNVPAVYSEKHEO
SIRONASAITLOWVYVAKDNGTSESETPAPRAEYKFKKDGALSIOPDNDNY
VYMDENYLNPAELKGRSSSAGIKYILDNDEPDLFTPHRIHQVYKCSLKNKSV
LAKYIKTLDPAELKGRSSSAGIKYILDNDEPDLFTPHRIHQVYKCSLKNKSV
FGKRLDVLIDHMYPEAOVGVRIKCPDENSTSDVAIARQARLTMDPYKTKQOG
QITAGENSIMINOMPEYELPLPNIKADIDKYPCTKLAITFEFDYGDHISGIALAD
VLGIFKGYGYMARMGDSGYAQAANYINYNVNGKSRGSGTCSAETDGVEMPY
ASIGEDDSTVHILLNRRNDKRAKELKNNNTVYTGELTGDSTISSQIRKGVYLS
NQNNITFIEPNLTVYHIVLTSSK"
misc_feature
725..2053
/note="Region: family 9 glycosyl hydrolase domain"
misc_feature
2054..2545
/note="type IIc cellulose binding site"
misc_feature
2627..3106
/note="type IIb cellulose binding site"
misc_feature
3224..4195
/note="duplicated type IIb cellulose binding site"
misc_feature
4298..5869
/note="Region: family 44 glycosyl hydrolase domain"
misc_feature
2537..2626
/note="PT-linker"
misc_feature
3107..3223
/note="PT-linker"
misc_feature
4196..4295
/note="PT-linker"
BASE COUNT 1926 a 1034 c 1620 g 1424 t 1 others
ORIGIN
Query Match 2.8%; Score 21; DB 1; Length 6005;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 41 AAATGATGAATTTTGTGGA 61
|||||
Db 2053 AAATGATGAATTTTGTGGA 2073
```

```
REFERENCE 1 (bases 1 to 6416)
AUTHORS Pajli,A., Petcaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,
Hu,Y.M. and Dalian,P.W.
TITLE Truncated cellulase composition
JOURNAL Patent: JP 1999221086-A 2 17-AUG-1999;
COMMENT CLARIANT INTERNATIONAL LTD
OS Unknown
PN JP 1999221086-A/2
PD 17-AUG-1999
PE 21-SEP-1998 JP 1998283606
PR 19-SEP-1997 US 08/932571
PI PAULI ANDERSON,PETRAER L BAKURUISUTO,ROY M DANIEL,PI
GURAHAMU K FARTON,
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONOTISU WILLIAM
PC C12N15/09,C1103/386,C12N1/21,C12N9/42/(C12N1/21,C12R1:19),PC
(C12N9/42,C12R1:19),C12N15/00
CC
FH Key Location/Qualifiers
FT source 1..6416
FT /organism="Unknown".
FEATURES
source location/Qualifiers
1..6416
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2068 a 1082 c 1689 g 1541 t 36 others
ORIGIN
Query Match 2.8%; Score 21; DB 6; Length 6416;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 41 AAATGATGAATTTTGTGGA 61
|||||
Db 2070 AAATGATGAATTTTGTGGA 2090
```

```
RESULT 40
LOCUS AX281444 6725 bp DNA Linear PAT 02-NOV-2001
DEFINITION Sequence 108 from Patent WO0177376.
ACCESSION AX281444
VERSION AX281444.1 GI:16608699
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with metastasis
JOURNAL Patent: WO 0177376-A 108 18-OCT-2001;
Epigenomics AG (DE)
FEATURES
source Location/Qualifiers
1..6725
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2074 a 53 c 1429 g 3169 t
ORIGIN
Query Match 2.8%; Score 21; DB 6; Length 6725;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 737 ATAAATTTTATTTTATTT 757
|||||
Db 1132 ATAAATTTTATTTTATTT 1112
```

```
RESULT 41
LOCUS AX346111 6725 bp DNA Linear PAT 01-FEB-2002
DEFINITION Sequence 1182 from Patent WO0200928.
ACCESSION AX346111
```

VERSION AX346111.1 GI:18493997
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1182 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES Location/Qualifiers
Source 1..6725
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2074 a 53 c 1429 g 3169 t
ORIGIN

Query Match 2.8%; Score 21; DB 6; Length 6725;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTATTT 757
|||||
Db 1132 ATAAATTTTATTTATTT 1112

RESULT 42
AX251852 9760 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 113 from Patent WO0168911.
ACCESSION AX251852
VERSION AX251852.1 GI:15985207
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 9760)
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the cell cycle
JOURNAL Patent: WO 0168911-A 113 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES Location/Qualifiers
Source 1..9760
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2848 a 292 c 2049 g 4571 t
ORIGIN

Query Match 2.8%; Score 21; DB 6; Length 9760;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTATTT 757
|||||
Db 9483 ATAAATTTTATTTATTT 9503

RESULT 43
AX344238 9760 bp DNA linear PAT 01-FEB-2002
LOCUS
DEFINITION Sequence 85 from Patent WO0200926.
ACCESSION AX344238
VERSION AX344238.1 GI:18492126
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with signal transduction
JOURNAL Patent: WO 0200926-A 85 03-JAN-2002;

FEATURES EpiGenomics AG (DE)
Source Location/Qualifiers
1..9760
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2848 a 292 c 2049 g 4571 t
ORIGIN

Query Match 2.8%; Score 21; DB 6; Length 9760;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTATTT 757
|||||
Db 9483 ATAAATTTTATTTATTT 9503

RESULT 44
AX348629 9760 bp DNA linear PAT 06-FEB-2002
LOCUS
DEFINITION Sequence 87 from Patent WO0202807.
ACCESSION AX348629
VERSION AX348629.1 GI:18614664
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with cell signalling
JOURNAL Patent: WO 0202807-A 87 10-JAN-2002;
EpiGenomics AG (DE)
FEATURES Location/Qualifiers
Source 1..9760
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2848 a 292 c 2049 g 4571 t
ORIGIN

Query Match 2.8%; Score 21; DB 6; Length 9760;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTATTT 757
|||||
Db 9483 ATAAATTTTATTTATTT 9503

RESULT 45
AF346733 15360 bp DNA linear INV 09-APR-2001
LOCUS
DEFINITION Tetrahymena thermophila cytoplasmic dynein 1 (DYH1) gene, partial cds.
ACCESSION AF346733
VERSION AF346733.1 GI:13561924
KEYWORDS
SOURCE Tetrahymena thermophila.
ORGANISM Tetrahymena thermophila.
REFERENCE 2 (bases 1 to 15360)
AUTHORS Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida; Tetrahymenina; Tetrahymena.
1 (bases 1 to 15360)
Demall,K.M., Wisniewski,J.C., Lee,S., Chang,H., Vanderpool,C. and Asai,D.J.
TITLE Cytoplasmic dynein 1 (DYH1) sequence from Tetrahymena thermophila
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 15360)
AUTHORS Demall,K.M., Wisniewski,J.C., Lee,S., Chang,H., Vanderpool,C. and Asai,D.J.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2001) Biological Sciences, Purdue University, Lilly Hall of Life Sciences, West Lafayette, IN 47907-1901, USA

Db 10861 ATAAATTTTATTTTATTT 10881

RESULT 47
AX348899 15954 bp DNA linear PAT 06-FEB-2002

LOCUS AX348899 Sequence 357 from Patent WO0202807.
ACCESSION AX348899
VERSION AX348899.1 GI:18614934

KEYWORDS
SOURCE . synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with cell signalling
JOURNAL Patent: WO 0202807-A 357 10-JAN-2002;
Epigenomics AG (DE)
FEATURES
source 1..15954
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4723 a 139 c 3285 g 7807 t

ORIGIN

Query Match 2.8%; Score 21; DB 6; Length 15954;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 737 ATAAATTTTATTTTATTT 757
Db 10861 ATAAATTTTATTTTATTT 10881
|||||
|||||

RESULT 48
AF036705/c 28606 bp DNA linear INV 22-MAY-2002

LOCUS AF036705/c
DEFINITION Caenorhabditis elegans cosmid F37C4, complete sequence.
ACCESSION AF036705
VERSION AF036705.2 GI:14625166
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderiinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 28606)
AUTHORS Waterston, R.
TITLE Genome sequence of the nematode C. elegans: a platform for
JOURNAL investigating biology. The C. elegans Sequencing Consortium
MEDLINE Science 282 (5396), 2012-2018 (1998)
PUBMED 99069613
9851916

REFERENCE 2 (bases 1 to 28606)
AUTHORS Langston, Y., Sansone, J. and Wohldmann, P.
TITLE The sequence of C. elegans cosmid F37C4
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 28606)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 28606)
Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
5 (bases 1 to 28606)
Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Louis, MO 63110, USA
6 (bases 1 to 28606)
Waterston, R.
Direct Submission
Submitted (22-MAY-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Jul 6, 2001 this sequence version replaced gi:2749976.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
email: rtwematode.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:
http://www.wormbase.org/db/seq/sequence?name=F37C4;class=Sequence

NEIGHBORING COSMID INFORMATION

NOTES:
The 5' cosmid is F28F9, 1100 bp overlap; the 3' cosmid is F56D6,
2400 bp overlap. Actual start of this cosmid is at base position
197 of F37C4; actual end is at 15237 of F56D6.

Coding sequences below are the result of integration and manual
review of the following data: computer analysis using the program
GeneFinder (P. Green and L. Miller, personal communication), the
large scale EST Projects of Yui Kohara
(http://www.ddb.jnc.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
elegans ORFome cloning project (http://wormfdb.fci.harvard.edu/),
similarity to other proteins from BlastX analyses
(http://blast.wustl.edu/), sequence conservation with C. briggsae
using Jim Kent's WABA alignment program (Genome Research
10:1115-1125, 2000), individual C. elegans Genbank submissions,
and personal communications with C. elegans researchers. tRNAs
are predicted using the program tRNAscan-SE (Lowe, T.M. and
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES
source
Location/Qualifiers
1..28606
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="F37C4"
3417..7733
/gene="F37C4.4"
/note="For a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=F37C4.4;class
=Sequence"
join(3417..3440,5558..5629,6271..6362,6649..6920,
7268..7446,7623..7733)
/gene="F37C4.4"
/note="coded for by the following C. elegans cDNAs:
YK45397.5"
/codon_start=1
/product="Hypothetical protein F37C4.4"
/protein_id="AAB95170.1"

gene	/codon_start=1 /product="Hypothetical protein F37C4.7" /protein_id="AAB95173.1" /db_xref="GI:2749983" /translation="MEIOPDIEKHAFVLRFPQPSIMVSLKIHIOFOHNSDGLIV RLSFAELAKPKIIIVISYOSPSQIEIAYDLVLKTKQRNEDRKQDDHMFSTILLIS LAICTLIIILSLFELR" complement(15324..17194) /gene="F37C4.8" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=F37C4.8;class =Sequence"	
CDS	complement(join(15324..15329,15378..15510,15596..15660, 16258..16497,16551..16845,16900..17071,17131..17194)) /gene="F37C4.8" /codon_start=1 /product="Hypothetical protein F37C4.8" /protein_id="AAB95174.1"	
Query Match	2.8%: Score 21; DB 3; Length 28606;	
Best Local Similarity	100.0%; Pred. No. 15;	
Matches 21; Conservative	0; Mismatches	0; Gaps 0;
QY 733 TAAATTAATTTTATTTT 753		
Db 1890 TAAATTAATTTTATTTT 1870		
RESULT 49		
AX409733	33206 bp	DNA linear PAT 14-JUN-2002
DEFINITION	Sequence 2380 from Patent WO0229103.	
ACCESSION	AX409733	
VERSION	AX409733.1	GI:21442438
KEYWORDS	human.	
SOURCE		
ORGANISM	Homo sapiens	
REFERENCE	1	
AUTHORS	Alvarres,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.	
TITLE	Gene expression profiles in liver cancer	
JOURNAL	Patent: WO 0229103-A 2380 11-APR-2002;	
FEATURES	GENE LOGIC INC (US)	
source	Location/Qualifiers	
	1..33206	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/note="EMBL/Genbank Accession No. M64554"	
BASE COUNT	9901 a 5689 c 5955 g 11661 t	
ORIGIN		
Query Match	2.8%: Score 21; DB 6; Length 33206;	
Best Local Similarity	100.0%; Pred. No. 14;	
Matches 21; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
QY 739 AAAATTTTATTTATTTCT 759		
Db 30390 AAAATTTTATTTATTTCT 30410		
RESULT 50		
HUMBFXIII	33206 bp	DNA linear PRI 31-OCT-1994
LOCUS	Human factor XIII b subunit gene, complete cds.	
DEFINITION	M64554 J05294	
ACCESSION	M64554.1	GI:179416
VERSION	blood coagulation factor; factor XIII; factor XIIIb; zymogen.	
KEYWORDS	Human DNA.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 33206)	

AUTHORS Botteanus,R.E., Ichinose,A. and Davie,E.W.
TITLE Nucleotide sequence of the gene for the b subunit of human factor XIII
JOURNAL Biochemistry 29, 11196-11209 (1990)
REFERENCE 2 (sites)
AUTHORS Nishimura,D.Y., Laysens,N.J. and Murray,J.C.
TITLE A dinucleotide repeat for the DIS3 locus
JOURNAL Nucleic Acids Res. 20 (5), 1167 (1992)
MEDLINE 92195854
PUBMED 1549502
FEATURES
source location/Qualifiers
1..33206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6p25-p24"
/cell_type="fibroblast; leukocyte"
/tissue_type="liver"
/dev_stage="fetus"
join(<2931..2994,7093..7293,8181..8366,9073..9249,9606..9782,12684..12863,12951..13136,14251..14433,17308..17508,19264..19446,29399..29612,30723..30943)/
/gene="f13a1"
join(<2931..2994,7093..7293,8181..8366,9073..9249,9606..9782,12684..12863,12951..13136,14251..14433,17308..17508,19264..19446,29399..29612,30723..30943)/
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
join(2931..2994,7093..7293,8181..8366,9073..9249,9606..9782,12684..12863,12951..13136,14251..14433,17308..17508,19264..19446,29399..29612,30723..30756)/
/gene="f13a1"
/codon_start=1
/product="coagulation factor XIIIb"
/protein_id="AAAS1821.1"
/db_xref="GI:179417"
/db_xref="GDB:G00-120-614"
/translation="MRLKNLFIILIIISGEIYAEEKPCGFPHVENGRIAOYYVFKS
FYFPMSIDKLSFPCLAGYTESGROEOTCTSGWSPEPRCFKCKTPDLISNGYIS
DVKLKYKIDENMHYCASGYKTTGGKDEYVQCLSDGSSOPTCKREHETCTAPELYN
GNYSTTOKTFKFKDYOEACGYTAGCKRTEVECLTYGMSLTPKCKIKLCCSIRL
IENGFEHPKQYEEGDVYQFCHENYIISGSDLIQCYRFGNYPESPVCEGRNNCP
PPLPINSKIOTHSITYRGEIVHIEELFELHGSAEIICEDEKSTPEPKIEGGEKV
ACEEPPFLENGAANLSKITYNGDKVYACKSGYLIHGSEITCNNGKWTLPCEVEN
NENCKRHPVYMNAGVADGIIASATGSSVEYKCNBYLIIRGSKIISRCDEGKSSPVC
LEPCIVNVDYNNRNINEMKMYEGKVLHGDLIDPCKOGYDLSPLTSELISVQCNRG
EVKYPLCTRKESKSGCTSPRLIKHGVIITISVDTYENGSSEYRCDFDHRFLGSRAY
CLDGKMTTPRLCEPCTLSFTMEKANLLIKWDFONRPHILHGEYIEFICRGDTTPAE
LITIGSILMOCDRQOLKIPRCIPROSTLSIODEPLKT"
<2931..2994
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=1
2995..7092
/gene="f13a1"
/note="G00-120-614"
/number=1
7093..7293
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=2
7294..8180
/gene="f13a1"
/note="G00-120-614"
/number=2
8181..8366
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=3
8367..9072

/gene="f13a1"
/note="G00-120-614"
/number=3
9073..9249
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=4
9250..9605
/gene="f13a1"
/note="G00-120-614"
/number=4
9606..9782
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=5
9783..12683
/gene="f13a1"
/note="G00-120-614"
/number=5
12684..12863
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=6
12864..12950
/gene="f13a1"
/note="G00-120-614"
/number=6
12951..13136
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=7
13137..14250
/gene="f13a1"
/note="G00-120-614"
/number=7
14251..14433
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=8
14434..17307
/gene="f13a1"
/note="G00-120-614"
/number=8
17308..17508
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=9
17509..19263
/gene="f13a1"
/note="G00-120-614"
/number=9
19264..19446
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=10
19447..29398
/gene="f13a1"
/note="G00-120-614"
/number=10
29399..29612
/gene="f13a1"
/product="coagulation factor XIIIb"
/note="G00-120-614"
/number=11
29613..30722
/gene="f13a1"

/note="G00-120-614"
/number=11
30723 . 30943

exon
/gene="F13A1"
/product="coagulation factor XIIB"
/note="G00-120-614"

polyA-signal
/number=12
30919 . 30924

/gene="F13A1"
/note="G00-120-614"

BASE COUNT 9901 a 5689 c 5955 g 11661 t
ORIGIN

Query Match 2.8%; Score 21; DB 9; Length 33206;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 739 AAAATTTTATTTTATTTCT 759
|||||

Db 30390 AAAATTTTATTTTATTTCT 30410

Search completed: February 20, 2003, 23:22:13
Job time : 4219 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 21:22:47 ; Search time 239 Seconds
(without alignments)
7170.589 Million cell updates/sec

Title: US-10-024-955-6
 Report: 761

Sequence: 1 GATCTTATATCATAACAAT.....ATTTTTATTTTATTTCTCC 761

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

word size :

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneset_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	761	100.0	612	15	AAO71401	House dust mite al
2	761	6.2	812	15	AAO71400	House dust mite al
3	22	24	15	AAO71405	Primer Df5 for hou	
4	22	3.2	11473	24	ABK40029	Human chemically p
5	22	2.9	11473	24	ABL33354	Human immune syste
6	21	2.8	28	15	AAO71406	Primer Df6 for hou
7	21	2.8	1442	22	AAO87276	Arabidopsis thalia
8	21	2.8	6435	20	AAK53662	DNA sequence encod
9	21	2.8	6416	24	AAO25626	Active cellulase p

C	10	21	2.8	6725	24	ABL33209	Human immune syste
C	11	21	2.8	6725	24	ABL34555	Human metastasis a
C	12	21	2.8	9760	24	ABL70197	Chemically treatee
C	13	21	2.8	9760	24	AS61152	Human gene regula
C	14	21	2.8	9760	24	ABK31242	Signal transductio
C	15	21	2.8	15954	24	ABL70467	Chemically treatee
C	16	21	2.8	15954	24	ABK31494	Signal transductio
C	17	21	2.8	33206	24	ABN95882	Gene #2380 used t
C	18	21	2.8	37973	24	ABL34197	Human immune syste
C	19	21	2.8	61020	22	AS46787	Tumour suppressor
C	20	20	2.6	794	22	AAH03680	Human cDNA clone
C	21	20	2.6	1107	22	AAH03617	Human nervous sys
C	22	20	2.6	1120	22	ABAI9618	Human nervous sys
C	23	20	2.6	1120	22	ABAI9619	Human nervous sys
C	24	20	2.6	2286	22	AAH17085	Human cDNA sequen
C	25	20	2.6	2783	23	ABL18600	Drosophila melano
C	26	20	2.6	4111	22	AAH76982	Human quinoprotei
C	27	20	2.6	4189	23	ABL07460	Drosophila melano
C	28	20	2.6	5910	24	ABL70177	Chemically treatee
C	29	20	2.6	5910	24	AS61130	Human gene regula
C	30	20	2.6	5910	24	ABK31218	Signal transductio
C	31	20	2.6	5928	24	ABL34086	Human immune syste
C	32	20	2.6	7403	22	AS456803	Tumour suppressor
C	33	20	2.6	7403	24	ABL34216	Human immune syste
C	34	20	2.6	8451	24	ABK39982	Human chemically E
C	35	20	2.6	8451	24	ABL32659	Human immune syste
C	36	20	2.6	8451	24	AS63319	Chemically pretres
C	37	20	2.6	9542	20	AAK20260	Borrelia burgdorfe
C	38	20	2.6	13420	24	ABL32917	Human immune syste
C	39	20	2.6	16217	24	ABL32625	Human immune syste
C	40	20	2.6	17185	22	ABAI6049	Human nervous syste
C	41	20	2.6	17185	22	AAK75628	Human immune/haem
C	42	19	2.5	333	14	AAO60657	Human brain Express
C	43	19	2.5	541	21	AAK07940	Fusarium venenatum
C	44	19	2.5	591	22	AAH55310	S. epidermidis ope
C	45	19	2.5	597	23	ABV51017	Human prostate exp

ALIGNMENTS

[illegible]

Query Match 6.2%; Score 47; DB 15; Length 812;
Best Local Similarity 100.0%; Pred. No. 6.4e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 GCCGTTTCGGCTGATTCACATGATGATAAAATCACCAGAAAT 128
|||||
DB 107 GCCGTTTCGGCTGATTCACATGATGATAAAATCACCAGAAAT 153

RESULT 3
AA071405
ID AA071405 standard; cDNA; 24 BP.
XX
AC AA071405;
XX
DT 01-APR-1995 (first entry)
XX
DE Primer Df5 for house dust mite allergen DerfVII cDNA.
XX
KM Primer: Df5; DNA sequencing; DerfVII allergen; anti-allergic;
XX allergy diagnosis; ss.
XX
OS Dermatophagoides farinae.
XX
PN MO9420614-A.
XX
PD 15-SEP-1994.
XX
PE 11-MAR-1994; 94MO-AU00117.
XX
PR 12-MAR-1993; 93US-0031141.
XX 22-JUN-1993; 93US-0081540.
XX
PA (CHIL-) INST CHILD HEALTH RES.
XX
PI Chua K, Thomas WR;
XX
PI WPI: 1994-303021/37.
XX
PT New nucleic acid encoding specific dust mite allergens - and
XX related vectors, transformed cells, peptides and antibodies,
XX useful for desensitisation and diagnosis.
XX
PS Example 5; Page 33; 67pp; English.
XX
CC The DNA sequencing primer Df5 is derived from nucleotides 559-582
XX of DerfVII (AA071401) and is used in the polymerase chain reaction
XX amplification and sequencing of a cDNA clone encoding DerfVII from a
XX phage lambda-gel1 cDNA library. DerfVII antigen is useful as an
XX anti-allergic reagent for treating sensitivity to house dust mite
XX allergens.
XX
SQ Sequence 24 BP; 5 A; 5 C; 4 G; 10 T; 0 other;

Query Match 3.2%; Score 24; DB 15; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 559 TCAATCTTGATCCCAATTTTGGC 582
|||||
DB 1 TCAATCTTGATCCCAATTTTGGC 24

RESULT 4
ABK40029
ID ABK40029 standard; DNA; 11473 BP.
XX
AC ABK40029;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #56 strand 1.
XX

KM Human: ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KM cytosinatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KM UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WC200202806-A2.
XX
PD 10-JAN-2002.
XX
PE 29-JUN-2001; 2001MO-EP07470.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
PI WPI: 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
XX useful for detecting cytosine methylation state of genes associated
XX with pharmacogenomics and for therapy of diseases e.g. cancer -
XX
PS Claim 1; SEQ ID NO 111; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at
CC least 18 bases in length of a segment of the chemically pretreated DNA
CC of genes associated with pharmacogenomics according to one of the
CC sequences of the genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1
CC (NM_000497), CYP3A3 (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2
CC (NM_001979), OCLN (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360),
CC MRP (NM_004996, NM_019900, NM_019901, NM_019902, NM_019862, NM_019898,
CC NM_019899) and their complementary sequences, or a sequence (S1) chosen
CC from 87 sequences and their complements. The chemical pretreatment
CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)
CC into uracils. Also included are an oligomer (II) in particular an
CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in
CC each case at least one base sequence having a length of 9 nucleotides
CC which hybridises to or is identical to a chemically pretreated DNA of
CC genes associated with pharmacogenomics and their complements, arranged in
CC an array for analysing diseases associated with the methylation state
CC (CpG) and/or detecting SNPs (single nucleotide polymorphisms)
CC of the 87 sequences. The oligomers may also be used as PCR primers.
CC The set of 87 nucleic acids and their complements is useful for diagnosis
CC and therapy of solid tumours and cancer. The present sequence
CC represents one the 87 DNA sequences or its complement.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 11473 BP; 2862 A; 260 C; 2742 G; 5609 T; 0 other;

Query Match 2.9%; Score 22; DB 24; Length 11473;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 736 AATTAATTTTATTTATTTATTT 757
|||||
DB 7672 AATTAATTTTATTTATTTATTT 7693

RESULT 5
ABL33354
ID ABL33354 standard; DNA; 11473 BP.
XX
AC ABL33354;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1327.
XX

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiaeriosclerotic; antianaemic; cytosatic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antirheumatic; antidiabetic; antiparasitic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 OS Homo sapiens.
 XX WO200200928-A2.
 PN 03-JAN-2002.
 PD 02-JUL-2001; 2001WO-EP07537.
 PE 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 PS Claim 1; SEQ ID NO 1327; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SQ Sequence 11473 BP; 2862 A; 260 C; 2742 G; 5609 T; 0 other;
 Query Match 2.9%; Score 22; DB 24; Length 11473;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 736 AATAAAATTTTATTTTATTT 757
 ||||||||||||||||||
 DB 7672 AATAAAATTTTATTTTATTT 7693
 RESULT 6
 AAQ71406/c
 ID AAQ71406 standard; CDNA; 28 BP.
 XX
 AC AAQ71406;
 XX
 DT 01-APR-1995 (first entry)
 XX
 DE Primer Df6 for house dust mite allergen DerfVII CDNA.
 XX
 KW Primer: Df6; DNA sequencing; DerfVII allergen; antiallergic;
 KW allergy diagnosis; ss.
 XX
 OS Dermatophagoides farinae.
 XX
 PN WO9420614-A.
 XX
 PD 15-SEP-1994.
 XX
 PF 11-MAR-1994; 94WO-AU00117.

XX 12-MAR-1993; 93US-0031141.
 PR 22-JUN-1993; 93US-0081540.
 XX
 PA (CHIL-) INST CHILD HEALTH RES.
 XX
 PI Chua K, Thomas WR;
 DR WPI; 1994-303021/37.
 XX
 PT New nucleic acid encoding specific dust mite allergens - and
 PT related vectors, transformed cells, peptides and antibodies,
 PT useful for desensitisation and diagnosis.
 XX
 PS Example 5; Page 33; 67pp; English.
 CC The DNA sequencing primer Df6 is derived from nucleotides 684-664
 CC of DerfVII (AAQ71401) and is used in the polymerase chain reaction
 CC amplification and sequencing of a cDNA clone encoding DerfVII from a
 CC phage lambda-gt11 cDNA library. DerfVII antigen is useful as an
 CC antiallergic reagent for treating sensitivity to house dust mite
 CC allergens.
 XX
 SQ Sequence 28 BP; 7 A; 6 C; 2 G; 13 T; 0 other;
 Query Match 2.8%; Score 21; DB 15; Length 28;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 664 CGTGATTGGAAAAAATTAA 684
 ||||||||||||||||||
 DB 28 CGTGATTGGAAAAAATTAA 8
 RESULT 7
 AAC87276
 ID AAC87276 standard; DNA; 1442 BP.
 XX
 AC AAC87276;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Arabidopsis thaliana seed-specific KNAT411 promoter fragment.
 XX
 KW KNAT411 promoter; seed-specific promoter; expression vector;
 KW embryo development; transgenic plant; germline modification;
 KW site-specific recombination system; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200068388-A1.
 PD 16-NOV-2000.
 XX
 PF 05-MAY-2000; 2000WO-EP04879.
 XX
 PR 06-MAY-1999; 99US-0306060.
 XX
 PA (RHOB-) RHOBIO.
 XX
 PI Terry LT, Hsieh T;
 DR WPI; 2001-007399/01.
 XX
 PT A seed-specific promoter from Arabidopsis, useful for controlling gene
 PT expression and activating recombination systems, active from an early
 PT stage of embryo development -
 XX
 PS Example 4; Fig 4A; 71pp; English.
 CC The invention relates to the Arabidopsis KNAT411 promoter (AAC87275)
 CC which directs seed-specific expression beginning in the early embryo.
 CC The invention also relates to expression cassettes and expression

CC vectors comprising the KNAT411 promoter operably linked to either a
CC heterologous gene, a native plant gene antisense sequence, or a Cre, FLP,
CC R or Gin recombinase gene: host plant cells transformed with a vector of
CC the invention; and transgenic plants, and seeds and progeny thereof,
CC regenerated from cells transformed with the vector. The KNAT411 promoter
CC is used in plant transformation vectors to direct seed-specific
CC expression of heterologous genes or antisense sequences targeted to an
CC endogenous plant gene, or to activate a site-specific recombination
CC system in the early embryo which results in a recombination event that
CC is fixed in the germline of the plant. In particular, the promoter is
CC used to increase or decrease (by antisense or co-suppression techniques)
CC levels of genes involved in fatty acid synthesis or lipid metabolism,
CC and an expression cassette of the invention is used to excise or invert
CC anonymous genes involved in embryo or seed development, or those of
CC unknown function, for which no stable mutations are available. The
CC KNAT411 promoter is active at a much earlier stage in embryo development
CC than known seed-specific promoters. The present sequence represents a
CC Arabidopsis KNAT411 promoter fragment.

XX Sequence 1442 BP; 516 A; 185 C; 199 G; 542 T; 0 other;

Query Match

Best Local Similarity 2.8%; Score 21; DB 22; Length 1442;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTATTT 757
DB 737 ATAAATTTTATTTATTT 757

RESULT 8

AAK55662
ID AAK55662 standard; DNA; 6415 BP.

AC AAK55662;

DT 30-JUL-1999 (first entry)

XX DNA sequence encoding truncated cellulases.

KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

XX cotton-containing fabric; stonewashing; ss.

OS Unidentified.

PN EP921188-A2.

PD 09-JUN-1999.

XX 15-SEP-1998; 98EP-0810919.

PF 19-SEP-1997; 97US-0932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

PA Anderson P, Bergquist PL, Daniels RM, Farrington GK;

PI Gibbs MD, Morgan H, Williams DP;

XX WPI: 1999-315403/27.

DR P-PsDB; AAY13493.

XX New truncated cellulase proteins, useful in detergents and for

PT producing 'stonewashed' denim

XX Disclosure; Page 24-25; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of

CC proteinases of native thermophilic and alkaliphilic origin, comprising

CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,

CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-

CC length sequences, or functional equivalents. Cel B5 extends from amino

CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino

CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
CC K1751 and the stability region extends from amino acid E482 to G635 in
CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or
CC remove staining, backstaining or graying, for use on cellulosic
CC materials including cotton-containing fabrics. They are especially useful
CC for preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using
CC non-truncated cellulase compositions.

XX Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 36 other;

Query Match

Best Local Similarity 2.8%; Score 21; DB 20; Length 6415;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 AAATGATGAATTTTGTGA 61

DB 2070 AAATGATGAATTTTGTGA 2090

RESULT 9

AAD26526
ID AAD26526 standard; DNA; 6416 BP.

AC AAD26526;

DT 26-MAR-2002 (first entry)

XX Active cellulase protein, cele gene.

KW Active cellulase protein; alkaliphilic; textile processing; proteinase;

KW detergent additive; stonewashed appearance; cotton-containing denim;

XX CelB5; thermophilic; commercial detergent; cele gene; ds.

OS Unidentified.

PN US6294366-B1.

PD 25-SEP-2001.

XX 19-AUG-1998; 98US-0136574.

PF 19-SEP-1997; 97US-0932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

PA Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;

PI Morgan H, Williams DP;

XX WPI: 2002-081780/11.

DR P-PsDB; AAE16324.

XX New cellulase active protein, useful in textile processing or

PT commercial detergents, e.g. for improving the feel or appearance of

PT cotton-containing fabrics, is stable under conditions of alkaline pH

XX and elevated temperatures -

XX Disclosure; Column 37-44; 61pp; English.

KW Human; signal transduction associated gene; cytosine methylation state;
 KW Cpg island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 PN WO200200926-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP07472.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-147896/19.
 XX
 PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction -
 PS
 PS Claim 1; SEQ ID NO 85; 24pp; English.
 XX
 CC The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphite or
 CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
 CC for detecting the cytosine methylation state (Cpg islands) of these
 CC genes, and a method for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with signal transduction.
 CC The genomic DNA can be obtained from cells or cellular components which
 CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
 CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
 CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
 CC histologic object slides, and all their possible combinations. The
 CC sequences of the invention are useful for the diagnosis and therapy of
 CC diseases associated with signal transduction e.g. solid tumours and
 CC cancer. ABR31158-ABR3155 represent chemically pretreated genomic DNA
 CC sequences of different genes associated with signal transduction, or
 CC their complementary sequences.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 9760 BP; 2848 A; 292 C; 2049 G; 4571 T; 0 other;
 Query Match 2.8%; Score 21; DB 24; Length 9760;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 737 ATAAAAATTTTATTTTATTT 757
 DB 9483 ATAAAAATTTTATTTTATTT 9503
 RESULT 15
 ABL70467
 ID ABL70467 standard; DNA; 15954 BP.
 XX
 AC ABL70467;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Chemically treated cell signalling DNA sequence#179.
 XX
 KW Cell signalling; cytosine methylation; cell signalling disease;
 KW cancer; tumour; cytostatic; ds.
 XX

OS unidentified.
 XX
 PN WO200202807-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP07471.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-154758/20.
 XX
 PT Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signalling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signalling -
 PS
 PS Claim 1; SEQ ID NO 357; 24pp+sequence listing; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signalling. The activity of the modified sequences of the
 CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signalling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signalling.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SQ Sequence 15954 BP; 4723 A; 139 C; 3285 G; 7807 T; 0 other;
 Query Match 2.8%; Score 21; DB 24; Length 15954;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 737 ATAAAAATTTTATTTTATTT 757
 DB 10861 ATAAAAATTTTATTTTATTT 10881
 RESULT 16
 ABR31494
 ID ABR31494 standard; DNA; 15954 BP.
 XX
 AC ABR31494;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified DNA #169.
 XX
 KW Human; signal transduction associated gene; cytosine methylation state;
 KW Cpg island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 PN WO200200926-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP07472.

DR WPI: 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PR for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1: SEQ ID NO 2170; 32pp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 37973 BP; 10756 A; 620 C; 8310 G; 18287 T; 0 other;
 Query Match 2.8%; Score 21; DB 24; Length 37973;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 737 ATAAATTTTATTATTATT 757
 DB 18993 ATAAATTTTATTATTATT 19013
 RESULT 19
 AAS46787/c
 ID AAS46787 standard; DNA; 61020 BP.
 XX
 AC AAS46787;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #513.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 OS
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2001-602752/68.
 XX
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer -
 XX
 PS Claim 1: SEQ ID NO 513; 27pp; English.
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since

CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with Cpg dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 61020 BP; 17256 A; 634 C; 13343 G; 29787 T; 0 other;
 Query Match 2.8%; Score 21; DB 22; Length 61020;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 737 ATAAATTTTATTATTATT 757
 DB 36205 ATAAATTTTATTATTATT 36185
 RESULT 20
 AAH03680
 ID AAH03680 standard; cDNA; 794 BP.
 XX
 AC AAH03680;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:515.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW
 OS Homo sapiens.
 OS
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 1: SEQ ID 515; 2537pp + CD ROM; English.
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH1629 to AAH1632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 794 BP; 268 A; 88 C; 136 G; 297 T; 5 other;

Query Match 2.6%; Score 20; DB 22; Length 794;

Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 740 AAATTTTATTTTATTTCT 759
|||||

Db 190 AAATTTTATTTTATTTCT 209

RESULT 21

ABAI9617 standard; DNA; 1107 BP.

XX ABA19617;

XX 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 11948.

XX Human: neotropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerrary;
KW antiparkinsonian; antischling; antianaemic; antitartaric; cancer;
KW antitumoural; hepatotoxic; cerebroprotective; antiinflammatory;
KW antifungal; antidiabetic; antidiabetic; antidiabetic; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225256.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235844.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.

08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-541565/50.
 DR
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Disclosure: SEQ ID NO 11948; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;

CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1107 BP; 326 A; 187 C; 200 G; 394 T; 0 other;
 Query Match 2.6%; Score 20; DB 22; Length 1107;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 738 TAAATTTTATTTTATTT 757
 DB 491 TAAATTTTATTTTATTT 510
 RESULT 22
 ABA19618
 ID ABA19618 standard; DNA; 1120 BP.
 XX
 AC ABA19618;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Human nervous system related polynucleotide SEQ ID NO 11949.
 XX
 KW Human; nootropic; neuroprotective; cytosolic; dermatological; virocidic;
 KW Immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antischistosomal; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200159063-A2.
 XX
 PD 16-AUG-2001.
 XX
 FE 17-JAN-2001; 2001WO-US01334.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EM;
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 7273; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL1840-ABL16175) and the encoded proteins
 CC (AB87737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 2783 BP; 801 A; 646 C; 595 G; 741 T; 0 other;
 XX
 QY Query Match 2.6%; Score 20; DB 23; Length 2783;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 DB 40 AAAATGATGAATTTTGT 59
 ||||||||||||||||
 DB 2165 AAAATGATGAATTTTGT 2184
 XX
 RESULT 26
 AAH76982
 ID AAH76982 standard; cDNA; 4111 BP.
 XX
 AC AAH76982;
 XX
 DT 15-DEC-2001 (first entry)
 XX
 DE Human quinoprotein reductase 13-encoding cDNA.
 XX
 KW Human; quinoprotein reductase 13; recombinant production;
 KW malignant tumour; cancer; blood disease; HIV infection;
 KW human immunodeficiency virus; immune disorder; inflammatory condition;
 KW cytosolic; anti-HIV; antiinflammatory; immunomodulator; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 3084..3446
 FT /*tag= a
 FT /product= "Human quinoprotein reductase 13"
 XX
 PN WO200173046-A1.

XX 04-OCT-2001.
 PD
 XX
 PF 26-MAR-2001; 2001WO-CN00452.
 XX
 PR 28-MAR-2000; 2000CN-0115253.
 XX
 PA (SHAN-) SHANGHAI BLOWINDOW GENE DEV INC.
 XX
 PI Mao Y, Xie Y;
 DR WPI; 2001-616513/71.
 DR P-PSDB; AAG66805.
 XX
 PT Human quinoprotein reductase 13 and encoded polynucleotide, applicable
 PT in diagnosis and treatment of malignant tumor, hemopathy, HIV
 PT infection, immunological diseases and inflammations -
 XX
 PS Claim 6; Page 27-29; 33pp; Chinese.
 XX
 CC This sequence represents cDNA encoding human quinoprotein reductase 13.
 CC The protein has a molecular weight of 13 kD. The invention relates to
 CC human quinoprotein reductase 13 (AAG66805), nucleic acids encoding it
 CC (AAH76982), and a method for the recombinant production of quinoprotein
 CC reductase 13. The present invention additionally discloses an antagonist
 CC of quinoprotein reductase 13 for therapeutic use, and an antibody which
 CC specifically binds to quinoprotein reductase 13. Quinoprotein reductase
 CC 13, and nucleotides which encode it may be used for treating a variety
 CC of diseases, such as malignant tumours, blood diseases, HIV (human
 CC immunodeficiency virus) infection, immune disorders and inflammatory
 CC conditions. The protein may also be used to screen for modulators of its
 CC activity or for peptide fingerprinting identification. The polynucleotide
 CC can be used as a primer for nucleic acid amplification reactions or as a
 CC probe for hybridisation reactions, or in producing gene chips or
 CC microarrays.
 XX
 SQ Sequence 4111 BP; 1297 A; 807 C; 706 G; 1301 T; 0 other;
 XX
 QY Query Match 2.6%; Score 20; DB 22; Length 4111;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 DB 738 TAAATTTTATTTTATTTT 757
 ||||||||||||||||
 DB 2318 TAAATTTTATTTTATTTT 2337
 XX
 RESULT 27
 ABL07460
 ID ABL07460 standard; cDNA; 4189 BP.
 XX
 AC ABL07460;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16862.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PsDB; ABB63357.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 16862; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB1840-AB16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX
 SQ Sequence 4189 BP; 1145 A; 917 C; 864 G; 1263 T; 0 other;
 XX
 Query Match 2.6%; Score 20; DB 23; Length 4189;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 OY 40 AAATGATGAATTTTGT 59
 |||||||
 Db 956 AAAATGATGAATTTTGT 975
 |||||||
 XX
 RESULT 28
 ABL70177
 ID ABL70177 standard; DNA: 5910 BP.
 XX
 AC ABL70177;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Chemically treated cell signalling DNA sequence#34.
 XX
 KW Cell signalling; cytosine methylation; cell signalling disease;
 KW cancer; tumour; cytostatic; ds.
 XX
 OS unidentified.
 XX
 PN WO200202807-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP07471.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-154758/20.
 XX
 PT Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signalling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signalling -
 XX
 PS Claim 1; SEQ ID NO 67; 24pp+sequence listing; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signalling. The activity of the modified sequences of the

CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signalling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signalling.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 CC
 XX
 SQ Sequence 5910 BP; 1457 A; 186 C; 1307 G; 2960 T; 0 other;
 XX
 Query Match 2.6%; Score 20; DB 24; Length 5910;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 OY 738 TAAATTTTATTTTATTT 757
 |||||||
 Db 846 TAAATTTTATTTTATTT 865
 |||||||
 XX
 RESULT 29
 AAS61130
 ID AAS61130 standard; DNA: 5910 BP.
 XX
 AC AAS61130;
 XX
 DT 29-JAN-2002 (first entry)
 XX
 DE Human gene regulation-associated gene oligonucleotide #85.
 XX
 KW Human: Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Hemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preeclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunostimulant; cardiant; antiinflammatory; coagulant; antisthmatic;
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200177375-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03968.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 XX
 PR 07-APR-2000; 2000DE-1019173.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-017470/02.
 XX
 PT New nucleic acid sequences from chemically modified genes associated
 PT with gene regulation, useful for analysing cytosine methylations for
 PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
 PT disease -
 XX
 PS Claim 1; SEQ ID NO 87; 26pp; English.
 XX
 CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The
 CC chemical pretreatment converts cytosine bases unmethylated at the

CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC precociousness, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct-sequences
XX
SQ Sequence 5910 BP; 1457 A; 186 C; 1307 G; 2960 T; 0 other;

Query Match 2.6%; Score 20; DB 24; Length 5910;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 738 TAAAAATTTTATTTTATTT 757
| | | | | | | | | | | | | | |
Db 846 TAAAAATTTTATTTTATTT 865

RESULT 30
ABK31218
ID ABK31218 standard; DNA; 5910 BP.
XX
AC ABK31218;
XX
XX 23-APR-2002 (first entry)
DE Signal transduction associated gene modified DNA #31.
XX
XX Human; signal transduction associated gene; cytosine methylation strate;
KW Cpg island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX WO200200926-A2.
PN
XX 03-JAN-2002.
PD
XX 29-JUN-2001; 2001WO-EP07472.
PF
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI: 2002-147896/19.
DR
XX
XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction -
XX
XX Claim 1; SEQ ID No 61; 24pp; English.
PS
XX
XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or pna oligomers
CC for detecting the cytosine methylation state (Cpg islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and

CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 5910 BP; 1457 A; 186 C; 1307 G; 2960 T; 0 other;

Query Match 2.6%; Score 20; DB 24; Length 5910;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 738 TAAAAATTTTATTTTATTT 757
| | | | | | | | | | | | | | |
Db 846 TAAAAATTTTATTTTATTT 865

RESULT 31
ABL34086
ID ABL34086 standard; DNA; 5928 BP.
XX
XX ABL34086;
XX
XX 26-MAR-2002 (first entry)
DE Human immune system associated gene SEQ ID NO: 2059.
XX
XX Human; immune system disease; cytosine methylation; antiautomatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootrophic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; Rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
PN
XX 03-JAN-2002.
PD
XX 02-JUL-2001; 2001WO-EP07537.
PF
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI: 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX Claim 1; SEQ ID NO 2059; 32pp + Sequence Listing; German.
PS
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SO Sequence 5928 BP; 1649 A; 110 C; 1223 G; 2946 T; 0 other;

Query Match 2.6%; Score 20; DB 24; Length 5928;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 734 AAAATAAATTTTATTTT 753
 ||||||||||||||||
 Db 2067 AAAATAAATTTTATTTT 2086

RESULT 32
 AAS46803/c
 ID AAS46803 standard; DNA; 7403 BP.

AC AAS46803;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #529.

DE Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.

OS Homo sapiens.

PN WO200168912-A2.

PD 20-SEP-2001.

PE 15-MAR-2001; 2001WO-EP02955.

PR 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2001-602752/68.

PS Claim 1; SEQ ID No 529; 27pp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with Cpg dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the

CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 7403 BP; 2500 A; 54 C; 1298 G; 3551 T; 0 other;

Query Match 2.6%; Score 20; DB 22; Length 7403;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 732 ATAAATAAATTTTATTT 751
 ||||||||||||||||
 Db 6044 ATAAATAAATTTTATTT 6025

RESULT 33
 ABL34216/c
 ID ABL34216 standard; DNA; 7403 BP.

AC ABL34216;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2189.

KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
 KW antirheumatic; antiarthritic; antidiabetic; antiparasitic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PE 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PS Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -

PS Claim 1; SEQ ID NO 2189; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

SO Sequence 7403 BP; 2500 A; 54 C; 1298 G; 3551 T; 0 other;

Query Match 2.6%; Score 20; DB 24; Length 7403;
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 732 ATAAATATAATTTTATTTT 751
|||||
Db 6044 ATAAATATAATTTTATTTT 6025

RESULT 34

ABK39982/c
ID ABK39982 standard; DNA; 8451 BP.

XX ABK39982;

XX 21-MAY-2002 (first entry)

DE Human chemically pretreated gene sequence #32 strand 2.

XX Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytosilatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

XX Homo sapiens.

XX MO200202806-A2.

XX 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP07470.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-154757/20.

PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated
PT with pharmacogenomics and for therapy of diseases e.g. cancer -

XX Claim 1; SEQ ID NO 64; 24pp; English.

XX The invention relates to a nucleic acid comprising a sequence at
CC least 18 bases in length of a segment of the chemically pretreated DNA
CC of genes associated with pharmacogenomics according to one of the
CC sequences of the genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1
CC (NM_000497), CYP3A3 (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2
CC (NM_001979), OCLN (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360),
CC MRP (NM_004996, NM_019900, NM_019901, NM_019902, NM_019862, NM_019898,
CC NM_019899) and their complementary sequences, or a sequence (SI) chosen
CC from 87 sequences and their complements. The chemical pretreatment
CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)
CC into uracils. Also included are an oligomer (II) in particular an
CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in
CC each case at least one base sequence having a length of 9 nucleotides
CC which hybridises to or is identical to a chemically pretreated DNA of
CC genes associated with pharmacogenomics and their complements, arranged in
CC an array for analysing diseases associated with the methylation state
CC (CpG) and/or detecting SNPs (single nucleotide polymorphisms)
CC of the 87 sequences. The oligomers may also be used as PCR primers.
CC The set of 87 nucleic acids and their complements is useful for diagnosis
CC and therapy of solid tumours and cancer. The present sequence
CC represents one the 87 DNA sequences or its complement.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 8451 BP; 2238 A; 127 C; 2012 G; 4074 T; 0 other;

Query Match 2.6%; Score 20; DB 24; Length 8451;

Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 25 AAAACATATCTTACAAAT 44
|||||
Db 5981 AAAACATATCTTACAAAT 5962

RESULT 35

ABL32659/c
ID ABL32659 standard; DNA; 8451 BP.

XX ABL32659;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 632.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosilatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -

XX Claim 1; SEQ ID NO 632; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/intestine bowel
CC diseases. The present sequence is a gene of the invention.

XX Sequence 8451 BP; 2238 A; 127 C; 2012 G; 4074 T; 0 other;

Query Match 2.6%; Score 20; DB 24; Length 8451;

Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 AAAACATATCTTACAAAT 44
|||||

Db 5981 AAAACATATCTTACAAAT 5962

RESULT 36

AAS63319/c
ID AAS63319 standard; DNA; 8451 BP.

XX	AA563319;
XX	
DT	29-JAN-2002 (first entry)
XX	
DE	Chemically pretreated metabolism associated gene #14.
XX	
KW	Human: cytosstatic; anti-tumour; metabolism; metabolic disease; liver;
KW	solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
KW	single nucleotide polymorphism detection; SNP; stool; urine; lung;
KW	cerebral-spinal fluid; intestines; brain; heart; prostate; breast;
KW	DUSP2; EPHX2; QDPR; GSGH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.
XX	
OS	Homo sapiens.
XX	
PN	MO200176451-A2.
XX	
PD	18-OCT-2001.
XX	
PF	06-APR-2001; 2001MO-EP04016.
XX	
PR	06-APR-2000; 2000DE-1019058.
XX	
PR	07-APR-2000; 2000DE-1019173.
XX	
PR	30-JUN-2000; 2000DE-1032529.
XX	
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIC-) EPIGENOMICS AC.
XX	
P1	Olek A, Piepenbrock C, Berlin K;
XX	
PI	WPI; 2002-010834/01.
XX	
PT	New nucleic acid, useful for diagnosis and therapy of metabolic
XX	disease, solid tumour and cancers, comprises segment of chemically
XX	modified genomic sequences of genes associated with metabolism -
XX	
PS	Claim 1; Page 54-56; 143pp; English.
XX	
CC	The invention relates to a nucleic acid (I) comprising a sequence at
CC	least 18 bases of a segment of the chemically pretreated DNA of genes
CC	associated with metabolism such as DUSP2 (NM_004418), EPHX2 (NM_001979),
CC	QDPR (NM_000320), GSGH (NM_000199), SHMT2 (NM_005412), SLC7A2
CC	(NM_003046), SLC7A4 (NM_004173) and TYMS (NM_001071) (all
CC	undefined). (I) are useful for diagnosis and therapy of metabolic
CC	disease, solid tumours and cancers; as primer oligonucleotides for the
CC	amplification of DNA sequences; for detecting the cytosine methylation
CC	state and/or single nucleotide polymorphisms (SNPs) in a chemically
CC	created DNA of genes associated with metabolism. An array of (I) is
CC	useful for ascertaining genetic and/or epigenetic parameters for the
CC	diagnosis and/or therapy of existing diseases or the predisposition to
CC	specific diseases by analysing cytosine methylations. The method involves
CC	chemically treating genomic DNA sample by a solution of bisulphite,
CC	hydrogen sulphite or disulphite such that cytosine bases which are
CC	unmethylated at the 5th-position are converted to uracil or another base
CC	which is dissimilar to cytosine in terms of hybridisation behaviour and
CC	amplifying fragments of the chemically pretreated genomic DNA. The
CC	genomic DNA is from cells or cellular components which contain DNA,
CC	sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,
CC	stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as
CC	tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast
CC	or liver, histologic object slides and their combinations. Genetic
CC	parameters are mutations, in particular insertions, deletions, point
CC	mutations, inversions and polymorphisms of genes associated with
CC	metabolism and sequences further required for their regulation.
CC	Epigenetic parameters are in particular cytosine methylations and
CC	further chemical modifications of DNA bases of genes associated with
CC	metabolism. Further epigenetic parameters include for e.g. the
CC	acetylation of histones which correlates with DNA methylation.
CC	AA563306-AA563313 represent chemically pretreated metabolism associated
CC	genes, and related primers of the invention.
XX	
90	Sequence 8451 BP; 2238 A; 127 G; 2012 G; 4074 T; 0 other;

Query Match	2.6%;	Score 20;	DB 24;	Length 8451;
Best Local Similarity	100.0%;	Pred. No. 11;		
Matches	20;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
OY	25	AAAAACATATCTTACAAAT	44	
Db	5981	AAAAACATATCTTACAAAT	5962	
RESULT 37				
ID	AAAX20260/c			
AC	AAAX20260	standard; DNA: 9542 BP.		
DT	AAAX20260;			
XX	04-MAY-1999	(first entry)		
DE	Borrelia burgdorferi	polynucleotide sequence #13.		
XX	Borrelia burgdorferi;	spirochete; bacterium; pathogen; Lyme disease;		
XX	epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;			
KW	infection; diagnosis; characterisation; detection; ds.			
XX	Borrelia burgdorferi.			
OS	W09858943-A1.			
XX	30-DEC-1998.			
XX	18-JUN-1998;	98MO-US12764.		
PF	03-SEP-1997;	97US-0057483.		
XX	20-JUN-1997;	97US-0050359.		
PR	22-JUL-1997;	97US-0053344.		
XX	22-JUL-1997;	97US-0053377.		
PA	(HUMA-) HUMAN GENOME SCI INC.			
PA	(MEDI-) MEDIMMUNE INC.			
XX	Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO;			
PI	White OR;			
XX	WPI: 1999-081217/07.			
DR	New isolated Borrelia burgdorferi nucleic acids - used to develop			
XX	products for the detection, diagnosis, characterisation, prevention			
PT	and therapy of infections, particularly Lyme disease			
XX	Claim 1: Page 920-925; 1128pp; English.			
XX	AAAX20248 to AAAX20402 represent polynucleotide sequences isolated from			
CC	Borrelia burgdorferi (Bb). Products derived from Bb can be used for			
CC	the detection, diagnosis, characterisation, prevention and therapy of			
CC	Bb infections, e.g. Lyme disease. They can also be used for the			
CC	production of biosynthetic products, e.g. enzymes. Borrelia belongs			
CC	to a family of mobile, spiral-shaped bacteria called Spirochetes.			
CC	Spirochetes are pathogenic in humans and Borrelia causes epidemic and			
CC	endemic relapsing fever, and Lyme borreliosis, more commonly known as			
CC	Lyme disease.			
XX	Sequence 9542 BP; 3812 A; 1160 C; 1113 G; 3457 T; 0 other:			
XX	Query Match	2.6%;	Score 20;	DB 20;
XX	Best Local Similarity	100.0%;	Pred. No. 11;	Length 9542;
XX	Matches	20;	Conservative	0;
XX			Mismatches	0;
XX			Indels	0;
XX			Gaps	0;
OY	734	AAAAATPAAATTTTATTTT	753	
Db	2055	AAAAATPAAATTTTATTTT	2036	
RESULT 38				
ABL32917				

```

ID ABL32917 standard; DNA; 13420 BP.
XX
AC ABL32917;
XX
DT 26-MAR-2002 (first entry)
DE
XX
DE Human immune system associated gene SEQ ID NO: 890.
XX
KW Human: immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosinetic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antineumatic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Plepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
PS
PS Claim 1; SEQ ID NO 890; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 13420 BP; 4031 A; 132 C; 2704 G; 6553 T; 0 other;
XX
Query Match 2.6%; Score 20; DB 24; Length 13420;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 735 AAAATATAATTTTATTTTAA 754
DB 4981 AAATATAATTTTATTTTAA 5000
IIIIIIIIIIIIIIIIIIII
RESULT 39
ABL32625/c
ID ABL32625 standard; DNA; 16217 BP.
XX
AC ABL32625;
XX
DT 26-MAR-2002 (first entry)
DE
XX
DE Human immune system associated gene SEQ ID NO: 598.
XX
KW Human: immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosinetic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

```

```

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Plepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
PS
PS Claim 1; SEQ ID NO 598; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 16217 BP; 5200 A; 131 C; 2955 G; 7930 T; 1 other;
XX
Query Match 2.6%; Score 20; DB 24; Length 16217;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 738 TAAATTTTATTTTATTTT 757
DB 10117 TAAATTTTATTTTATTTT 10098
IIIIIIIIIIIIIIIIIIII
RESULT 40
ABA16049
ID ABA16049 standard; DNA; 17185 BP.
XX
AC ABA16049;
XX
DT 23-JAN-2002 (first entry)
DE
XX
DE Human nervous system related polynucleotide SEQ ID NO 8380.
XX
KW Human: noctropic; neuroprotective; cytosinetic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antischizoid; antihaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX

```

PF 17-JAN-2001: 2001WO-US01314.
XX
PR 31-JAN-2000: 2000US-0179065.
PR 04-FEB-2000: 2000US-0180628.
PR 24-FEB-2000: 2000US-0184664.
PR 02-MAR-2000: 2000US-0186350.
PR 16-MAR-2000: 2000US-0189874.
PR 17-MAR-2000: 2000US-0190076.
PR 18-APR-2000: 2000US-0198123.
PR 19-MAY-2000: 2000US-0205515.
PR 07-JUN-2000: 2000US-0209467.
PR 28-JUN-2000: 2000US-0214886.
PR 30-JUN-2000: 2000US-0215135.
PR 07-JUL-2000: 2000US-0216647.
PR 07-JUL-2000: 2000US-0216680.
PR 11-JUL-2000: 2000US-0217487.
PR 11-JUL-2000: 2000US-0217496.
PR 14-JUL-2000: 2000US-0218290.
PR 26-JUL-2000: 2000US-0220963.
PR 26-JUL-2000: 2000US-0220964.
PR 14-AUG-2000: 2000US-0224518.
PR 14-AUG-2000: 2000US-0224519.
PR 14-AUG-2000: 2000US-0225213.
PR 14-AUG-2000: 2000US-0225214.
PR 14-AUG-2000: 2000US-0225266.
PR 14-AUG-2000: 2000US-0225267.
PR 14-AUG-2000: 2000US-0225268.
PR 14-AUG-2000: 2000US-0225270.
PR 14-AUG-2000: 2000US-0225447.
PR 14-AUG-2000: 2000US-0225757.
PR 14-AUG-2000: 2000US-0225758.
PR 14-AUG-2000: 2000US-0225759.
PR 18-AUG-2000: 2000US-0226279.
PR 22-AUG-2000: 2000US-0226681.
PR 22-AUG-2000: 2000US-0226688.
PR 22-AUG-2000: 2000US-0227182.
PR 23-AUG-2000: 2000US-0227009.
PR 30-AUG-2000: 2000US-0228924.
PR 01-SEP-2000: 2000US-0229287.
PR 01-SEP-2000: 2000US-0229343.
PR 01-SEP-2000: 2000US-0229344.
PR 01-SEP-2000: 2000US-0229345.
PR 05-SEP-2000: 2000US-0229509.
PR 05-SEP-2000: 2000US-0229513.
PR 06-SEP-2000: 2000US-0230437.
PR 06-SEP-2000: 2000US-0230438.
PR 08-SEP-2000: 2000US-0231242.
PR 08-SEP-2000: 2000US-0231243.
PR 08-SEP-2000: 2000US-0231244.
PR 08-SEP-2000: 2000US-0231413.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0232080.
PR 08-SEP-2000: 2000US-0232081.
PR 12-SEP-2000: 2000US-0231968.
PR 14-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232398.
PR 14-SEP-2000: 2000US-0232399.
PR 14-SEP-2000: 2000US-0232400.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0233063.
PR 14-SEP-2000: 2000US-0233064.
PR 14-SEP-2000: 2000US-0233065.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234997.
PR 25-SEP-2000: 2000US-0234998.
PR 26-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235836.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236367.
PR 29-SEP-2000: 2000US-0236368.
PR 29-SEP-2000: 2000US-0236369.

PR 29-SEP-2000: 2000US-0236370.
PR 02-OCT-2000: 2000US-0236802.
PR 02-OCT-2000: 2000US-0237037.
PR 02-OCT-2000: 2000US-0237038.
PR 02-OCT-2000: 2000US-0237039.
PR 02-OCT-2000: 2000US-0237040.
PR 13-OCT-2000: 2000US-0239935.
PR 13-OCT-2000: 2000US-0239937.
PR 20-OCT-2000: 2000US-0240960.
PR 20-OCT-2000: 2000US-0241785.
PR 20-OCT-2000: 2000US-0241786.
PR 20-OCT-2000: 2000US-0241787.
PR 20-OCT-2000: 2000US-0241808.
PR 20-OCT-2000: 2000US-0241809.
PR 20-OCT-2000: 2000US-0241826.
PR 20-OCT-2000: 2000US-0242221.
PR 01-NOV-2000: 2000US-0244617.
PR 08-NOV-2000: 2000US-0246474.
PR 08-NOV-2000: 2000US-0246475.
PR 08-NOV-2000: 2000US-0246476.
PR 08-NOV-2000: 2000US-0246477.
PR 08-NOV-2000: 2000US-0246478.
PR 08-NOV-2000: 2000US-0246523.
PR 08-NOV-2000: 2000US-0246524.
PR 08-NOV-2000: 2000US-0246525.
PR 08-NOV-2000: 2000US-0246526.
PR 08-NOV-2000: 2000US-0246527.
PR 08-NOV-2000: 2000US-0246528.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246509.
PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249267.
PR 17-NOV-2000: 2000US-0249299.
PR 17-NOV-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250391.
PR 01-DEC-2000: 2000US-0251160.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251868.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251889.
PR 08-DEC-2000: 2000US-0251989.
PR 11-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PR 05-JAN-2001: 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM:
XX
DR WPI: 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,

PT useful for preventing, diagnosing and/or treating nervous system
cancers and metastases -
XX
PS
XX Disclosure: SEQ ID NO 8380; 1701bp + Sequence Listing: English.
CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABAI678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (anti)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcc_sequences.

PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Disclosure: SEQ ID NO 30440; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703

CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SO Sequence 17185 BP; 3967 A; 4732 C; 4679 G; 3807 T; 0 other;
 Query Match 2.6%; Score 20; DB 22; Length 17185;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 738 TAAATTTTATTTTATTT 757
 Db 14538 TAAATTTTATTTTATTT 14519
 RESULT 42
 AAO60657 standard; CDNA; 333 BP.
 XX
 AC AAO60657;
 XX
 DT 16-MAR-1994 (first entry)
 XX
 XX Human brain Expressed Sequence Tag EST02672.
 DE
 KW Gene transcription product; genetic markers; tagging; in vivo;
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9316178-A.
 XX
 PD 19-AUG-1993.
 XX
 PF 12-FEB-1993; 93WO-US01294.
 XX
 PR 12-FEB-1992; 92US-0837195.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX
 PI Adams MD, Moreno RF, Venter CJ;
 XX
 DR WPI; 1993-272882/34.
 XX
 PT Enriched oligonucleotides and corresp. sequences - used as
 PT markers for human genes transcribed in-vivo, facilitate tagging
 PT of most human genes
 XX
 PS Example 4; Page 369; 500pp; English.
 XX
 CC The Expressed Sequence Tag was isolated from a human brain CDNA
 CC library as part of a large set of ESTs which can be used as markers
 CC for human genes transcribed in vivo. They can be used to facilitate
 CC tagging of most human genes, for mapping locations of expressed genes
 CC on chromosomes, for individual or forensic identification, for mapping
 CC locations of disease-associated genes, for identification of tissue
 CC type, and for prep. of antisense sequences, probes and constructs.
 CC EST02672 has a "poor" coding probability as evaluated using the
 CC coding-region Prediction program CRM. See also AAO59041-061440.
 XX
 SO Sequence 333 BP; 90 A; 64 C; 65 G; 114 T; 0 other;
 QY
 Query Match 2.5%; Score 19; DB 14; Length 333;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 739 AAAATTTTATTTTATTT 757
 Db 167 AAAATTTTATTTTATTT 185
 RESULT 43
 AAF07940

ID AAF07940 standard; cDNA; 541 BP.
 XX AAF07940;
 AC
 XX
 DT 13-MAR-2001 (first entry)
 XX
 XX Fusarium venenatum EST SEQ ID NO:463.
 DE
 XX Multiple gene expression: filamentous fungal cell; EST;
 XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 PD 28-SEP-2000.
 XX
 XX 22-MAR-2000; 2000WO-US07781.
 PF
 XX 22-MAR-1999; 99US-0273623.
 PR
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI; 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 562; 3161pp; English.
 XX
 XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organization of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 541 BP; 146 A; 149 C; 120 G; 126 T; 0 other;
 Query Match 2.5%; Score 19; DB 21; Length 541;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 175 ACAATGATCATGAAG 193
 DB 373 ACAATGATCATGAAG 391

RESULT 44
 ID AAF53310 standard; DNA; 591 BP.
 XX AAF53310
 AC
 XX
 DT 03-SEP-2001 (first entry)
 XX
 XX S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2013.
 DE
 XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 KW
 XX Staphylococcus epidermidis.
 OS
 XX
 PN WO200134809-A2.
 PD 17-MAY-2001.
 XX
 XX 09-NOV-2000; 2000WO-US30782.
 PF
 XX 09-NOV-1999; 99US-0164258.
 PR
 XX (GLAX) GLAXO GROUP LTD.
 PA
 XX
 PI Kimerly WJ;
 XX
 DR WPI; 2001-316495/33.
 DR P-PSDB; AAG82460.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 PT
 PS Claim 8; Page 546; 2188pp; English.
 XX
 XX AAF52304 to AAF53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAF53971 to
 CC AAF55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAF55091 to
 CC AAF55098 represent oligonucleotide sequences and primers which are used
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present invention,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 591 BP; 220 A; 88 C; 93 G; 190 T; 0 other;
 Query Match 2.5%; Score 19; DB 22; Length 591;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 384 GGAATGATTTAGCATAC 402
 DB 153 GGAATGATTTAGCATAC 171
 RESULT 45
 ID ABV51017/c standard; cDNA; 597 BP.
 XX ABV51017;
 AC
 XX

DT 17-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 51008.
DE
XX
XX Human: prostate cancer: cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
OS
XX Homo sapiens.
PN WO200160860-A2.
PD
XX 23-AUG-2001.
PF
XX 20-FEB-2001; 2001WO-US05171.
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
DR
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 9919; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 597 BP; 215 A; 117 C; 126 G; 139 T; 0 other:
Query Match 2.5%; Score 19; DB 23; Length 597;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 740 AAATTTTATTTTATTC 758
|||||
DB 209 AAATTTTATTTTATTC 191
RESULT 46
ABN92019
ID ABN92019 standard; DNA; 606 BP.
XX
AC ABN92019;
XX
XX 24-JUL-2002 (first entry)
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1482.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy; gene; ds.

XX
XX Staphylococcus epidermidis.
OS
XX US6380370-B1.
PN
XX 30-APR-2002.
PD
XX
XX 13-AUG-1998; 98US-0134001.
PE
XX
XX 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Doucette-Stamm LA, Bush D;
PI WPI; 2002-381255/41.
XX
DR P-PSDB; ABP39474.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 1482; 267pp; English.
PS
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 606 BP; 228 A; 84 C; 99 G; 195 T; 0 other:
Query Match 2.5%; Score 19; DB 24; Length 606;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 384 GGAATGATGATTACATAC 402
|||||
DB 168 GGAATGATGATTACATAC 186
RESULT 47
AAD00890
ID AAD00890 standard; cDNA; 2141 BP.
XX
AC AAD00890;
XX
XX 08-SEP-2000 (first entry)
DE Fusarium venenatum lysophospholipase encoding cDNA.
XX
XX Lysophospholipase; phospholipid content; edible oil; hydrolysis; dough;
KW aqueous carbohydrate solution; baked product; antistaling; ss.
XX
XX Fusarium venenatum.
OS
XX
FH Key Location/Qualifiers
FT CDS 100..2064
FT /*tag= a
FT /*product= "Fusarium venenatum lysophospholipase"
FT sig_peptide 100..150
FT /*tag= b
FT misc_feature 151..210
FT /*tag= c
FT /*label= Pro-sequence
FT mat_peptide 211..2061
FT /*tag= d

```
FT /product= "Mature Fusarium venenatum lysophospholipase"  
FT /function= "Hydrolyses the fatty acyl group(s) of a  
FT phospholipid or lysophospholipid"  
PN WO200028044-A1.  
XX  
PD 18-MAY-2000.  
XX  
PF 10-NOV-1999; 99WO-US26789.  
XX  
PR 10-NOV-1998; 98US-0189486.  
XX  
PA (NOVO ) NOVO NORDISK BIOTECH INC.  
PA (NOVO ) NOVO NORDISK BIOINDUSTRY LTD.  
PI Berka RM, Rey MM, Byun T, Itami R, Tsutsumi N, Klotz A;  
XX  
XX WPI: 2000-376556/32.  
DR P-PSDB; AA771236.  
XX  
PT Novel lysophospholipase polypeptides useful for reducing the  
PT phospholipid content in an edible oil, aqueous carbohydrate solution or  
PT slurry and preparing dough and baked products with improved properties  
PT  
PS Claim 18; Fig 1; 93pp; English.  
XX  
XX The present sequence is the Fusarium venenatum lysophospholipase encoding  
CC cDNA, isolated from mycelial samples of Fusarium venenatum CCl-3, a  
CC morphological mutant of Fusarium strain ATCC 20334. It shares identity  
CC with lysophospholipase proteins from Neurospora crassa, Penicillium  
CC notatum, Saccharomyces cerevisiae and Schizosaccharomyces pombe, between  
CC regions that are important for catalytic and/or structural roles of the  
CC enzyme. Lysophospholipase treatment is useful for reducing the  
CC phospholipid content in an edible oil, aqueous carbohydrate solution or  
CC slurry, by hydrolysis of the phospholipid. It is useful for preparing  
CC dough with improved properties like, increased strength, stability,  
CC reduced stickiness and improved machinability and for preparation of a  
CC baked product with increased volume, improved crumb structure, softness,  
CC flavour and antistaling.  
XX  
SQ Sequence 2141 BP; 548 A; 550 C; 481 G; 562 T; 0 other;  
Query Match 2.5%; Score 19; DB 21; Length 2141;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 175 ACAATGATCCCAATGAAG 193  
DB 343 ACAATGATCCCAATGAAG 361  
|||||  
RESULT 48  
AAH70175/C  
ID AAH70175 standard; DNA; 3366 BP.  
XX  
AC AAH70175;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:308.  
XX  
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO200025728-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-US26796.  
XX
```

```
PR 05-NOV-1998; 98US-0107131.  
XX  
PA (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX  
XX WPI: 2000-365347/31.  
DR  
XX  
PT Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX  
PS Disclosure; Page 509-510; 577pp; English.  
XX  
XX The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAH70078 to AAH70287 and ABH8144 to ABH18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 3366 BP; 1495 A; 229 C; 399 G; 1243 T; 0 other;  
Query Match 2.5%; Score 19; DB 21; Length 3366;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 731 AATTAATATAAATTTTAT 749  
DB 1217 AATTAATATAAATTTTAT 1199  
|||||  
RESULT 49  
AAH54895/C  
ID AAH54895 standard; DNA; 3585 BP.  
XX  
AC AAH54895;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4259.  
XX  
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KW vaccination; endocarditis; ds.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN WO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX
```

PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimmery MJ;
XX
DR WPI: 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1984-1985; 2188bp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 3585 BP; 1286 A; 583 C; 435 G; 1281 T; 0 other;
XX
Query Match 2.5%; Score 19; DB 22; Length 3585;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 384 GGAATGATGATTAGCATAC 402
DB 1491 GGAATGATGATTAGCATAC 1473
XXXXXXXXXXXXXXXXXXXX
RESULT 50
ABL33705
ID ABL33705 standard; DNA: 5759 BP.
XX
AC ABL33705;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene. SEQ ID NO: 1678.
XX
KW Human; Immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriasis;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX

PA (EPIC-) EPIDEMIOLOGICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 1678; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5759 BP; 1537 A; 58 C; 1220 G; 2944 T; 0 other;
XX
Query Match 2.5%; Score 19; DB 24; Length 5759;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 739 AAAATTTTATTATTATTATT 757
DB 4724 AAAATTTTATTATTATTATT 4742
XXXXXXXXXXXXXXXXXXXX

Search completed: February 20, 2003, 22:15:25
Job time : 464 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 22:15:32 : Search time 81 Seconds
(Without alignments)
2881.247 Million cell updates/sec

Title: US-10-024-955-6
Perfect score: 761
Sequence: 1 GATCTATATCATACAT.....ATTTCCTTCCTCC 761

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PCBUS.COMB.seq:*
6: /cgn2.6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	761	100.0	761	4 US-08-553-336A-6	Sequence 6, Appl
2	47	6.2	812	3 US-08-462-778-1	Sequence 1, Appl
3	47	6.2	812	4 US-08-553-336A-1	Sequence 1, Appl
4	24	3.2	24	4 US-08-553-336A-12	Sequence 12, Appl
5	21	2.8	28	4 US-08-553-336A-13	Sequence 13, Appl
6	21	2.8	6416	4 US-09-136-574A-2	Sequence 2, Appl
7	19	2.5	606	4 US-09-134-001C-1482	Sequence 1482, Ap
8	18	2.4	18	3 US-08-462-778-3	Sequence 3, Appl
9	18	2.4	18	4 US-08-553-336A-3	Sequence 3, Appl
10	18	2.4	18	4 US-08-553-336A-15	Sequence 15, Appl
11	18	2.4	26	4 US-08-553-336A-8	Sequence 8, Appl
12	17	2.2	426	4 US-09-328-111-304	Sequence 304, App
13	17	2.2	845	2 US-08-743-637B-25	Sequence 25, Appl
14	17	2.2	845	3 US-08-526-840B-25	Sequence 25, Appl
15	17	2.2	1381	2 US-08-454-557C-49	Sequence 49, Appl
16	17	2.2	1381	2 US-08-340-426D-49	Sequence 49, Appl
17	17	2.2	1381	2 US-08-450-673C-49	Sequence 49, Appl
18	17	2.2	1381	5 PCT-US95-17111A-49	Sequence 49, Appl
19	17	2.2	1418	5 PCT-US95-17111A-120	Sequence 120, App
20	17	2.2	1442	2 US-08-454-557C-120	Sequence 120, App
21	17	2.2	1442	2 US-08-340-426D-120	Sequence 120, App
22	17	2.2	1442	2 US-08-450-673C-120	Sequence 120, App
23	17	2.2	1674	1 US-08-365-981-2	Sequence 2, Appl
24	17	2.2	1679	4 US-09-306-060-1	Sequence 1, Appl
25	17	2.2	3396	4 US-08-974-549A-641	Sequence 641, App
26	17	2.2	3440	1 US-08-471-791-27	Sequence 27, Appl
27	17	2.2	3440	5 PCT-US91-01746-27	Sequence 27, Appl

C 28	17	2.2	3984	4 US-08-961-527-176	Sequence 176, App
C 29	17	2.2	4332	3 US-08-728-603-14	Sequence 14, Appl
C 30	17	2.2	7042	4 US-09-092-508-1	Sequence 1, Appl
C 31	17	2.2	7042	4 US-09-435-115-1	Sequence 1, Appl
C 32	17	2.2	7042	4 US-09-098-310-1	Sequence 1, Appl
C 33	17	2.2	7042	4 US-09-690-364-21	Sequence 21, Appl
C 34	17	2.2	7075	4 US-09-092-508-15	Sequence 15, Appl
C 35	17	2.2	7075	4 US-09-435-115-15	Sequence 15, Appl
C 36	17	2.2	8585	1 US-08-030-096-3	Sequence 3, Appl
C 37	17	2.2	9461	4 US-09-221-017B-513	Sequence 513, App
C 38	17	2.2	17327	1 US-07-906-871-15	Sequence 15, Appl
C 39	17	2.2	32207	2 US-08-770-379-20	Sequence 20, Appl
C 40	17	2.2	32207	4 US-08-757-669A-20	Sequence 20, Appl
C 41	17	2.2	32207	4 US-09-230-371A-20	Sequence 20, Appl
C 42	17	2.2	34185	4 US-09-545-481-3	Sequence 3, Appl
C 43	17	2.2	162450	4 US-09-345-882-1	Sequence 1, Appl
C 44	16	2.1	155	1 US-08-710-082-11	Sequence 11, Appl
C 45	16	2.1	155	3 US-08-913-462-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-553-336A-6
Sequence 6, Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS: 15
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEO ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 761 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 43..681
US-08-553-336A-6
Query Match 100.0%; Score 761; DB 4; Length 761;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 GATCTTATATCAATAACAATCCAAAAACATATCTTACAAAATGATGAATTTTGTG 60
    |||||||
DB 1 GATCTTATATCAATAACAATCCAAAAACATATCTTACAAAATGATGAATTTTGTG 60
OY 61 ATTGCTGCCGTGGCATTTTCCCGCTTGGCTGATCCAAATTCACATATATAAATCACC 120
    |||||||
DB 61 ATTGCTGCCGTGGCATTTTCCCGCTTGGCTGATCCAAATTCACATATATAAATCACC 120
OY 121 GAAGAAATCAACAAAGCTATGATGATGCCATTTGCTATTTGAACAAATCCGAACATA 180
    |||||||
DB 121 GAAGAAATCAACAAAGCTATGATGATGCCATTTGCTATTTGAACAAATCCGAACATA 180
OY 181 GATCCAAATGAAGTACCTGATGATGCCATTTGCTATTTGAACAAATCCGAACATA 240
    |||||||
DB 181 GATCCAAATGAAGTACCTGATGATGCCATTTGCTATTTGAACAAATCCGAACATA 240
OY 241 TTCAAAGTGAATTTAGCCATGCGAAACATTTGAGCTCGAGATTTGAACAAATGAACGT 300
    |||||||
DB 241 TTCAAAGTGAATTTAGCCATGCGAAACATTTGAGCTCGAGATTTGAACAAATGAACGT 300
OY 301 CAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
    |||||||
DB 301 CAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
OY 361 GTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
    |||||||
DB 361 GTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
OY 421 CCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
    |||||||
DB 421 CCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
OY 481 TCTGATGAAGTGAACATTAACAATGACATCTTTGAAGTGAAGTGAAGTGAAGTGAAGT 540
    |||||||
DB 481 TCTGATGAAGTGAACATTAACAATGACATCTTTGAAGTGAAGTGAAGTGAAGTGAAGT 540
OY 541 AACCATATTTGGTGGCTTTTCAATCTTGATGCCAATTTTGGCGTTTATCTGATGATG 600
    |||||||
DB 541 AACCATATTTGGTGGCTTTTCAATCTTGATGCCAATTTTGGCGTTTATCTGATGATG 600
OY 601 ACCGCTATTTCCAGACACCGCTAGTAAGAAATGCCAAATTTGGCAGCAGATTT 660
    |||||||
DB 601 ACCGCTATTTCCAGACACCGCTAGTAAGAAATGCCAAATTTGGCAGCAGATTT 660
OY 661 AAACGTGAATGGAATAAATTAACCAATAGACATCTTTTCCAACTGATCAATCTCTA 720
    |||||||
DB 661 AAACGTGAATGGAATAAATTAACCAATAGACATCTTTTCCAACTGATCAATCTCTA 720
OY 721 TTTCACTGACATTAATAAATTTTATTTTATTTCTCC 761
    |||||||
DB 721 TTTCACTGACATTAATAAATTTTATTTTATTTCTCC 761

```

```

RESULT 2
US-08-462-778-1
; Sequence 1, Application US/08462278
; Patent No. 6077517

```

```

; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kaw-Yan
; TITLE OF INVENTION: Allergenic Protein and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,778
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USN 08/031,141
; FILING DATE: 12 March 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 812 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 68..712
; US-08-462-778-1

Query Match 6.2%; Score 47; DB 3; Length 812;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 GCCGTTTCGGTGAATCCATTCATGATGAATAATCACCAGAGAAT 128
    |||||||
DB 107 GCCGTTTCGGTGAATTCATGATGAATAATCACCAGAGAAT 153

RESULT 3
US-08-553-336A-1
; Sequence 1, Application US/0855336A
; Patent No. 6413738
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214

```



```

CORRESPONDENCE ADDRESS:
ADDRESS: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Query Match          2.8%; Score 21; DB 4; Length 6416;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 AATGATGAAATTTTGTGA 61
|||||
DB 2070 AATGATGAAATTTGTGA 2090

RESULT 7
US-09-134-001C-1482
Sequence 1482, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1482
LENGTH: 606
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1482

Query Match          2.5%; Score 19; DB 4; Length 606;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 384 GGAATATGATTAGCATAC 402
|||||
DB 168 GGAATATGATTAGCATAC 186
```

```

RESULT 8
US-08-462-778-3
Sequence 3, Application US/08462778
Patent No. 6077517
GENERAL INFORMATION:
APPLICANT: Thomas, Wayne R.
APPLICANT: Chua, Kaw-Yan
TITLE OF INVENTION: Allergenic Protein and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Lahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,778
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/031,141
FILING DATE: 12 March 1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-462-778-3

Query Match          2.4%; Score 18; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 GATCCATTCACATGAT 111
|||||
DB 1 GATCCATTCACATGAT 18

RESULT 9
US-08-553-336A-3
Sequence 3, Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
```


MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-553-336A-3

Query Match 2.4%; Score 18; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GATCCAATTCACATGAT 111
Db 1 GATCCAATTCACATGAT 18

RESULT 10
US-08-553-336A-15/C
Sequence 15 Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:

LENGTH: 18 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-553-336A-15

Query Match 2.4%; Score 18; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 208 GATAAATTCGAACGTCAT 225
Db 18 GATAAATTCGAACGTCAT 1

RESULT 11
US-08-553-336A-8
Sequence 8 Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-553-336A-8

Query Match 2.4%; Score 18; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GATCCAATTCACATGAT 111
Db 9 GATCCAATTCACATGAT 26

RESULT 12
US-09-328-111-304/C
Sequence 304 Application US/09328111
Patent No. 6262333

```
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 304
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-328-111-304

Query Match      2.2%; Score 17; DB 4; Length 426;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 736 AATAAATTTTATT 752
    |||||||
Db 175 AATAAATTTTATT 159
```

```
RESULT 13
; US-08-743-637B-25/C
; Sequence 25, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: Roy, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
```

```
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Haemophilus influenzae
; US-08-743-637B-25
```

```
Query Match      2.2%; Score 17; DB 2; Length 845;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 734 AATAAATTTTATT 750
    |||||||
Db 616 AATAAATTTTATT 600
```

```
RESULT 14
; US-08-526-840B-25/C
; Sequence 25, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: Roy, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Haemophilus influenzae
; US-08-526-840B-25
```

Query Match 2.2%; Score 17; DB 3; Length 845;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 734 AAAATAAATTTTATT 750
Db 616 AAAATAAATTTTATT 600

RESULT 15

US-08-454-557C-49
Sequence 49, Application US/08454557C
Patent No. 5830670
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,557C
FILING DATE: 30-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609,3840003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 1381 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-454-557C-49

Query Match 2.2%; Score 17; DB 2; Length 1381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 741 AATTTTATTTTATT 757
Db 597 AATTTTATTTTATT 613

RESULT 16

US-08-340-426D-49
Sequence 49, Application US/08340426D
Patent No. 5948634
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600

CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,426D
FILING DATE: 14-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609,3840002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 1381 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-340-426D-49

Query Match 2.2%; Score 17; DB 2; Length 1381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 741 AATTTTATTTTATT 757
Db 597 AATTTTATTTTATT 613

RESULT 17

US-08-450-673C-49
Sequence 49, Application US/08450673C
Patent No. 5948888
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,673C
FILING DATE: 30-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609,3840004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 1381 base pairs

TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-450-673C-49

Query Match 2.2%; Score 17; DB 2; Length 1381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
|||||
DB 597 AATTTTATTTTATTT 613

RESULT 18
PCT-US95-17111A-49
; Sequence 49, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neutral Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1381 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
PCT-US95-17111A-49

Query Match 2.2%; Score 17; DB 5; Length 1381;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
|||||
DB 597 AATTTTATTTTATTT 613

RESULT 19
PCT-US95-17111A-120
; Sequence 120, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neutral Thread Protein Gene Expression and

; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE: CDS
; NAME/KEY: CDS
; LOCATION: 14..1418
PCT-US95-17111A-120

Query Match 2.2%; Score 17; DB 5; Length 1418;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
|||||
DB 592 AATTTTATTTTATTT 608

RESULT 20
US-08-454-557C-120
; Sequence 120, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neutral Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995

```
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609.3840003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 1442 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 15..1139
US-08-454-557C-120

Query Match          2.2%; Score 17; DB 2; Length 1442;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 741 AATTTTATTTTATTT 757
Db 593 AATTTTATTTTATTT 609

RESULT 21
US-08-340-426D-120
Sequence 120, Application US/08340426D
Patent No. 5948634
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
APPLICANT: Wands, Jack R.
TITLE OF INVENTION: Neutral Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340.426D
FILING DATE: 14-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609.3840002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 1442 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 15..1139
US-08-340-426D-120

Query Match          2.2%; Score 17; DB 2; Length 1442;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 741 AATTTTATTTTATTT 757
Db 593 AATTTTATTTTATTT 609

RESULT 22
US-08-450-673C-120
Sequence 120, Application US/08450673C
Patent No. 5948888
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
APPLICANT: Wands, Jack R.
TITLE OF INVENTION: Neutral Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450.673C
FILING DATE: 30-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609.3840004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 1442 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 15..1139
US-08-450-673C-120

Query Match          2.2%; Score 17; DB 2; Length 1442;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 741 AATTTTATTTTATTT 757
Db 593 AATTTTATTTTATTT 609

RESULT 23
US-08-365-981-2
Sequence 2, Application US/08365981
Patent No. 5583030
GENERAL INFORMATION:
APPLICANT: Robert DICKSON et al
APPLICANT: METHOD FOR OBTAINING
TITLE OF INVENTION: ANTIFUNGAL AND HERBICIDAL COMPOUNDS THAT TARGET THE
TITLE OF INVENTION: FIRST COMMITTED STEP IN SPHINGOLIPID LONG-CHAIN
```

;; TITLE OF INVENTION: BASE BIOSYNTHESIS
;; NUMBER OF SEQUENCES: 15
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER
;; STREET: STE. 300, 99 CANAL CENTER PLAZA
;; CITY: Alexandria
;; STATE: Virginia
;; COUNTRY: USA
;; ZIP: 22314
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: DOS Text File
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/365,981
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/906,899
;; FILING DATE: 06/30/92
;; ATTORNEY/AGENT INFORMATION:
;; NAME: D.J. Mills
;; REGISTRATION NUMBER: 34506
;; REFERENCE/DOCKET NUMBER: 434-028
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 684-1111
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1674
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: polypeptide
US-08-365-981-2

Query Match 2.28; Score 17; DB 1; Length 1674;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 383 TGGAAATGATGATTGCA 399
Db 584 TGGAAATGATGATTGCA 600

RESULT 24
US-09-306-060-1
;; Sequence 1, Application US/09306060
;; Patent No. 6342637
;; GENERAL INFORMATION:
;; APPLICANT: Thomas, Terry L.
;; APPLICANT: Fu-Hsih, Tzung
;; TITLE OF INVENTION: SEED SPECIFIC PROMOTERS
;; FILE REFERENCE: RHONE POULENC
;; CURRENT APPLICATION NUMBER: US/09/306,060
;; CURRENT FILING DATE: 1999-05-06
;; NUMBER OF SEQ ID NOS: 1
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1679
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
US-09-306-060-1

Query Match 2.28; Score 17; DB 4; Length 1679;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 741 AATTTTATTTTATTT 757
Db 880 AATTTTATTTTATTT 896

RESULT 25
US-08-974-549A-641
;; Sequence 641, Application US/08974549A
;; Patent No. 6166178
;; GENERAL INFORMATION:
;; APPLICANT: Cech, Thomas R.
;; APPLICANT: Lingner, Joachim
;; APPLICANT: Nakamura, Toru
;; APPLICANT: Chapman, Karen B.
;; APPLICANT: Morin, Gregg B.
;; APPLICANT: Harley, Calvin B.
;; APPLICANT: Andrews, William H.
;; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
;; NUMBER OF SEQUENCES: 727
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Townsend and Townsend and Crew LLP
;; STREET: Two Embarcadero Center, Eighth Floor
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-3834
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/974,549A
;; FILING DATE: 19-NOV-1997
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/724,643
;; FILING DATE: 01-OCT-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/844,419
;; FILING DATE: 18-APR-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/846,017
;; FILING DATE: 25-APR-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/851,843
;; FILING DATE: 06-MAY-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/854,050
;; FILING DATE: 09-MAY-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/911,312
;; FILING DATE: 14-AUG-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/912,951
;; FILING DATE: 14-AUG-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/915,503
;; FILING DATE: 14-AUG-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US97/17618
;; FILING DATE: 01-OCT-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US97/17685
;; FILING DATE: 01-OCT-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Apple, Randolph Ted
;; REGISTRATION NUMBER: 36,429
;; REFERENCE/DOCKET NUMBER: 015389-002610US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 641:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3396 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear

MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1, 3396
OTHER INFORMATION: /note="inrt-encoding sequence employing
alternative codon distributions for
OTHER INFORMATION: yeast (high expressing genes)"
US-08-974-549A-641

Query Match 2.2%; Score 17; DB 4; Length 3396;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 448 CAAGATTGTGTGTC 464
DB 3006 CAAGATTGTGTGTC 3022

RESULT 26

US-08-471-791-27
Sequence 27, Application US/08471791
Patent No. 5723595
GENERAL INFORMATION:
APPLICANT: Thompson, Gregory A
APPLICANT: Knaut, Vic C
TITLE OF INVENTION: Plant Desaturases-Compositions
TITLE OF INVENTION: and Uses
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: California
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: MicrosoftWord 5.1 (a)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,791
FILING DATE: 6-JUNE-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/762,762
FILING DATE: 16-SEPT-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/01746
FILING DATE: 14-MAR-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/615,784
FILING DATE: 14-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/567,373
FILING DATE: 13-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/494,106
FILING DATE: 16-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Lassen, Elizabeth
REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
NAME: Carl J. Schwedler
REGISTRATION NUMBER: 36,924
REFERENCE/DOCKET NUMBER: CGNE 69-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
TELEX: 350370 CGNE

INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3440 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-471-791-27

Query Match 2.2%; Score 17; DB 1; Length 3440;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 738 TAAATTTTATTTTA 754
DB 2352 TAAATTTTATTTTA 2368

RESULT 27

PCT-US91-01746-27
Sequence 27, Application PC/TUS9101746
GENERAL INFORMATION:
APPLICANT: Thompson, Gregory A
APPLICANT: Knaut, Vic C
TITLE OF INVENTION: Plant Desaturases-Compositions and Uses
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: California
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.7
SOFTWARE: MicrosoftWord 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/01746
FILING DATE: 19910314
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/615,784
FILING DATE: 14-NOV-1990
APPLICATION NUMBER: 07/567,373
FILING DATE: 13-AUG-1990
APPLICATION NUMBER: 07/494,106
FILING DATE: 16-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Lassen, Elizabeth
REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
REFERENCE/DOCKET NUMBER: CGNE 69-3 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
TELEX: 350370 CGNE
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3440 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
PCT-US91-01746-27

Query Match 2.2%; Score 17; DB 5; Length 3440;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 738 TAAATTTTATTTTA 754

Db 2352 TAAATTTTATTTTA 2368

RESULT 28
US-08-961-527-176/c

Sequence 176, Application US/08961527

Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 176:

SEQUENCE CHARACTERISTICS:

LENGTH: 3984 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-176

Query Match 2.2%; Score 17; DB 4; Length 3984;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 734 AAAATAAATTTTAT 750

Db 1177 AAAATAAATTTTAT 1161

RESULT 29

US-08-728-603-14/c

Sequence 14, Application US/08728603

Patent No. 6093806

GENERAL INFORMATION:

APPLICANT: Cesarman, Ethel

APPLICANT: Knowles, Daniel M.

TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED

TITLE OF INVENTION: HERPESVIRUS

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP

STREET: Clinton Square, P.O. Box 1051

CITY: Rochester

STATE: New York

COUNTRY: USA

ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/728,603

FILING DATE: 10-OCT-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BRAMAN, SUSAN J.

REGISTRATION NUMBER: 34,103

REFERENCE/DOCKET NUMBER: 19603/720

TELEPHONE: 716-263-1636

TELEFAX: 716-263-1600

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 4332 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-728-603-14

Query Match 2.2%; Score 17; DB 3; Length 4332;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 GTTGTGCTTGCCCT 473

Db 2902 GTTGTGCTTGCCCT 2886

RESULT 30

US-09-092-508-1

Sequence 1, Application US/09092508

Patent No. 6291643

GENERAL INFORMATION:

APPLICANT: Henzel, William J.

TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF CASPASE-3

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt

STREET: 3100 No. 6291643west Center, 90 South seventh st

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/092,508

FILING DATE: 05-JUN-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/048,807

FILING DATE: 05-JUN-1997

APPLICATION NUMBER: 60/055,258

FILING DATE: 07-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: Kettelberger, Ph.D., Denise M

REGISTRATION NUMBER: 33,924

REFERENCE/DOCKET NUMBER: 11669.6US01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7042 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 578...4159
OTHER INFORMATION:
US-09-092-508-1

Query Match 2.2%; Score 17; DB 4; Length 7042;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 666 TGAATTGCAAAAAAATT 682
|||||
DB 4202 TGAATTGCAAAAAAATT 4218

RESULT 31

US-09-435-115-1
Sequence 1, Application US/09435115
Patent No. 6346607

GENERAL INFORMATION:
APPLICANT: Henzel, William J.
TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 3100 No. 6346607 West Center, 90 South Seventh St
CITY: Minneapolis
STATE: MN

COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/435.115

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/092.508

FILING DATE: 60/055.258

ATTORNEY/AGENT INFORMATION:

NAME: Kettelberger, Ph.D., Denise M

REGISTRATION NUMBER: 33,924

REFERENCE/DOCKET NUMBER: 11669.60501

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7042 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 578...4159

OTHER INFORMATION:

US-09-435-115-1

Query Match 2.2%; Score 17; DB 4; Length 7042;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 666 TGAATTGCAAAAAAATT 682
|||||
DB 4202 TGAATTGCAAAAAAATT 4218

RESULT 32
US-09-098-310-1
Sequence 1, Application US/09098310
Patent No. 6403765

GENERAL INFORMATION:

APPLICANT: Aienm1, Emad S.

TITLE OF INVENTION: TRUNCATED APAF-1 AND METHODS OF USE

TITLE OF INVENTION: THEREFOR

FILE REFERENCE: 480140.438

CURRENT APPLICATION NUMBER: US/09/098.310

CURRENT FILING DATE: 1998-06-16

NUMBER OF SEQ ID NOS: 15

SOFTWARE: FASTSEQ for Windows Version 3.0

SEQ ID NO 1

LENGTH: 7042

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: CDS

LOCATION: (578)...(4159)

US-09-098-310-1

Query Match 2.2%; Score 17; DB 4; Length 7042;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 666 TGAATTGCAAAAAAATT 682
|||||
DB 4202 TGAATTGCAAAAAAATT 4218

RESULT 33

US-09-690-364-21

Sequence 21, Application US/09690364
Patent No. 6468795

GENERAL INFORMATION:

APPLICANT: Hong Zhang

TITLE OF INVENTION: ANTISENSE MODULATION OF APAF-1 EXPRESSION

FILE REFERENCE: PRTS-0190

CURRENT APPLICATION NUMBER: US/09/690.364

CURRENT FILING DATE: 2000-10-17

NUMBER OF SEQ ID NOS: 100

SEQ ID NO 21

LENGTH: 7042

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (578)...(4162)

US-09-690-364-21

Query Match 2.2%; Score 17; DB 4; Length 7042;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 666 TGAATTGCAAAAAAATT 682
|||||
DB 4202 TGAATTGCAAAAAAATT 4218

RESULT 34

US-09-092-508-15

Sequence 15, Application US/09092508
Patent No. 6291643

GENERAL INFORMATION:

APPLICANT: Henzel, William J.

TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3

```

; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6291643west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092, 508
; FILING DATE: 05-JUN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048, 807
; FILING DATE: 05-JUN-1997
; APPLICATION NUMBER: 60/055, 258
; FILING DATE: 07-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Ph.D., Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11669.6US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7075 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 578...4192
; OTHER INFORMATION:
;
US-09-092-508-15

Query Match          2.2%; Score 17; DB 4; Length 7075;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 666 TGAATTGCAAAAAAATT 682
DB 4235 TGAATTGCAAAAAAATT 4251

RESULT 35
US-09-435-115-15
; Sequence 15, Application US/09435115
; Patent No. 6346607
; GENERAL INFORMATION:
; APPLICANT: Henzel, William J.
; TITLE OF INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6346607west Center, 90 South Seventh St
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/09/435, 115
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/092, 508
; FILING DATE:
; APPLICATION NUMBER: 60/055, 258
; FILING DATE: 07-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Ph.D., Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 11669.6US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7075 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 578...4192
; OTHER INFORMATION:
;
US-09-435-115-15

Query Match          2.2%; Score 17; DB 4; Length 7075;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 666 TGAATTGCAAAAAAATT 682
DB 4235 TGAATTGCAAAAAAATT 4251

RESULT 36
US-08-030-096-3/C
; Sequence 3, Application US/08030096
; Patent No. 5426041
; GENERAL INFORMATION:
; APPLICANT: Fadljanski, Steven F.
; APPLICANT: Arnison, Paul G.
; TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID
; TITLE OF INVENTION: SEED PRODUCTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/030, 096
; FILING DATE: 22-MAR-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/556, 917
; FILING DATE: 20-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA91/00255
; FILING DATE: 22-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/164/PIHT
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8585 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(308..370, 1136..1261, 6369..6428, 7198..7353)
US-08-030-096-3

Query Match 2.2%; Score 17; DB 1; Length 8585;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
|||||
DB 4316 AATTTTATTTTATTT 4300

RESULT 37
US-09-221-017B-513/C
Sequence 513, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2811
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 513:
SEQUENCE CHARACTERISTICS:
LENGTH: 9461 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...9461
US-09-221-017B-513

Query Match 2.2%; Score 17; DB 4; Length 9461;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 711 ACAATCTCTATTTCAC 727
|||||
DB 751 ACAATCTCTATTTCAC 735

RESULT 38
US-07-906-871-15
Sequence 15, Application US/07906871
Patent No. 5340739
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1225 Connecticut Avenue, N.W., Suite 300
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906.871
FILING DATE: 19920103
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/816.289
FILING DATE: 03 JAN 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/535.544
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/03051
FILING DATE: 13-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/224.035
FILING DATE: 13-JUL-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cimbalia, Michele A
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0627.2830004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)833-7533
TELEFAX: (202)833-8716
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 17327 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: both
TOPOLOGY: linear

MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: exon
LOCATION: 621..753
FEATURE:
NAME/KEY: Intron
LOCATION: 754..9596
FEATURE:
NAME/KEY: exon
LOCATION: 9597..9744
FEATURE:
NAME/KEY: Intron
LOCATION: 9745..16396
FEATURE:
NAME/KEY: exon
LOCATION: 16397..17327
US-07-906-871-15

Query Match 2.2%; Score 17; DB 1; Length 17327;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTATTTTATTT 757
Db 7408 AATTTATTTTATTT 7424

RESULT 39
US-08-770-379-20
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 2.2%; Score 17; DB 2; Length 32207;

Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 457 GTTGTGCCCTGTGCCCT 473
Db 26321 GTTGTGCCCTGTGCCCT 26337

RESULT 40
US-08-757-669A-20
Sequence 20, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match 2.2%; Score 17; DB 4; Length 32207;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 GTTGTGCCCTGTGCCCT 473
Db 26321 GTTGTGCCCTGTGCCCT 26337

RESULT 41
US-09-230-371A-20
Sequence 20, Application US/09230371A
Patent No. 6348586
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND

```

; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
; US-09-230-371A-20

Query Match
Best Local Similarity 100.0%; DB 4; Length 32207;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 457 GTTGTGCTTCCTTCCT 473
Db 26321 GTTGTGCTTCCTTCCT 26337

RESULT 42
US-09-545-481-3
; Sequence 3, Application US/09545481
; Patent No. 6451319
; GENERAL INFORMATION:
; APPLICANT: Chiang, Christina H.
; APPLICANT: Cochran, Mark D.
; TITLE OF INVENTION: No. 6451319el Recombinant And Mutant Adenoviruses
; FILE REFERENCE: SY0993K US
; CURRENT APPLICATION NUMBER: US/09/545,481
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,766
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 34185
; TYPE: DNA
; ORGANISM: Bovine adenovirus type 1
; US-09-545-481-3

Query Match
Best Local Similarity 100.0%; DB 4; Length 34185;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 GCATTTGTCGCCGTTTC 89
Db 5312 GCATTTGTCGCCGTTTC 5328

RESULT 43
US-09-345-882-1
; Sequence 1, Application US/09345882
; Patent No. 6393373
; GENERAL INFORMATION:
; APPLICANT: Bouguetel, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 162450
; TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72794
; OTHER INFORMATION: 5-124-273 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88073
; OTHER INFORMATION: 5-127-261 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90842
; OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93714
; OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97122
; OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97152
; OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99098
; OTHER INFORMATION: 5-130-257 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99117
; OTHER INFORMATION: 5-130-276 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 103806
; OTHER INFORMATION: 5-131-395 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 106940
; OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108106
; OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108149
; OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108308
; OTHER INFORMATION: 5-135-357 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108471
; OTHER INFORMATION: 5-136-174 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134124
; OTHER INFORMATION: 5-140-120 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134362
; OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134374
; OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
; FEATURE:
; NAME/KEY: allele
; LOCATION: 146328
; OTHER INFORMATION: 5-143-84 : polymorphic base A or G
```

```

FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:

```

```

1 NAME/KEY: allele
2 LOCATION: 99094..99140
3 OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
4 FEATURE:
5 NAME/KEY: allele
6 LOCATION: 103783..103828
7 OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
8 FEATURE:
9 NAME/KEY: allele
10 LOCATION: 103783..103828
11 OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
12 FEATURE:
13 NAME/KEY: allele
14 LOCATION: 106918..106966
15 OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
16 FEATURE:
17 NAME/KEY: allele
18 LOCATION: 106918..106966
19 OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
20 FEATURE:
21 NAME/KEY: allele
22 LOCATION: 108084..108130
23 OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
24 FEATURE:
25 NAME/KEY: allele
26 LOCATION: 108127..108177
27 OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
28 FEATURE:
29 NAME/KEY: allele
30 LOCATION: 108127..108177
31 OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
32 FEATURE:
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
```

APPLICATION NUMBER: US/08/710,082
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carly, Christine E.
REGISTRATION NUMBER: 36,099
REFERENCE/DOCKET NUMBER: 194301A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)-594-6734
TELEFAX: (908)-594-4720
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-710-082-11

Query Match 2.1%; Score 16; DB 1; Length 155;
Best Local Similarity 100.0%; Pred. NO. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 CTGATCAATTCACCTA 107
|||||
DB 147 CTGATCAATTCACCTA 132

RESULT 45
US-08-913-462-11/c
Sequence 11, Application US/08913462
Patent No. 6159729
GENERAL INFORMATION:
APPLICANT: Holmann, Kathryn J.
APPLICANT: Neeper, Michael P.
TITLE OF INVENTION: RECOMBINANT HUMAN PAPILLOMAVIRUS TYPE 11
TITLE OF INVENTION: L1 PROTEIN
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dr. Christine E. Carly - Merck & Co., Inc.
STREET: 126 E. Lincoln Avenue - P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,462
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/413,572
FILING DATE: 30-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Carly, Christine E.
REGISTRATION NUMBER: 36,099
REFERENCE/DOCKET NUMBER: 19429
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-6734
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-913-462-11

Query Match 2.1%; Score 16; DB 3; Length 155;
Best Local Similarity 100.0%; Pred. NO. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 CTGATCAATTCACCTA 107
|||||
DB 147 CTGATCAATTCACCTA 132

RESULT 46
US-09-624-482-11/c
Sequence 11, Application US/09624482
Patent No. 6358744
GENERAL INFORMATION:
APPLICANT: Volkin, David B.
APPLICANT: Sanyal, Gautam
APPLICANT: Shi, Li
TITLE OF INVENTION: STABILIZED HUMAN PAPILLOMAVIRUS
TITLE OF INVENTION: FORMULATIONS
FILE REFERENCE: 19933Y
CURRENT APPLICATION NUMBER: US/09/624,482
CURRENT FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 09/056,067
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: 60/042,808
PRIOR FILING DATE: 1997-04-08
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 155
TYPE: DNA
ORGANISM: Synthetic DNA Oligomer
US-09-624-482-11

Query Match 2.1%; Score 16; DB 4; Length 155;
Best Local Similarity 100.0%; Pred. NO. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 CTGATCAATTCACCTA 107
|||||
DB 147 CTGATCAATTCACCTA 132

RESULT 47
US-08-687-080-93
Sequence 93, Application US/08687080
Patent No. 5965427
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.

```

;
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 5' END OF INTRON 16 OF RAD50 GENOMIC
; INDIVIDUAL ISOLATE: SEQUENCE
US-08-687-080-93

Query Match
Best Local Similarity 100.0%; Score 16; DB 2; Length 239;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 ATTTTATTTTATTT 757
DB 47 ATTTTATTTTATTT 62

RESULT 48
US-08-557-309B-4
; Sequence 4, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-557-309B-4

Query Match
Best Local Similarity 100.0%; Score 16; DB 2; Length 373;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 742 ATTTTATTTTATTT 757
DB 175 ATTTTATTTTATTT 190

RESULT 49
US-08-834-306-4
; Sequence 4, Application US/08834306
; Patent No. 6054135
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,306
; FILING DATE: 15-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-834-306-4

Query Match
Best Local Similarity 100.0%; Score 16; DB 3; Length 373;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 ATTTTATTTTATTT 757
DB 175 ATTTTATTTTATTT 190

RESULT 50
US-08-993-674A-4
; Sequence 4, Application US/08993674A
; Patent No. 6228372
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
```


CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentp Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 373 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-993-674A-4

Query Match 2.1%; Score 16; DB 4; Length 373;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0;

Qy 742 ATTTTATTTTATTT 757
|||||
Db 175 ATTTTATTTTATTT 190

Search completed: February 20, 2003, 23:18:51
Job time : 308 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 23:18:57 : Search time 114 Seconds
(without alignments)
3399.959 Million cell updates/sec

Title: US-10-024-955-6

Perfect score: 761

Sequence: 1 GATCTATATCATACATACAT.....ATTTCATTTTATTTCTCC 761

Scoring table: OLIGO_NMC

Gapop 60.0, Gapext 60.0

Searched: 424239 seqs, 254661826 residues

Word size: 0

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCrUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	761	100.0	761	9 US-10-024-955-6	Sequence 6, Appli
2	47	6.2	812	9 US-10-024-955-1	Sequence 12, Appli
3	24	3.2	24	9 US-10-024-955-12	Sequence 1, Appli
4	21	2.8	28	9 US-10-024-955-13	Sequence 13, Appli
5	21	2.8	33206	10 US-09-880-107-2360	Sequence 2380, Ap
6	7	2.5	1966	9 US-09-938-842A-3199	Sequence 3199, Ap
7	19	2.5	513509	9 US-09-754-853A-4	Sequence 4, Appli
8	9	2.5	640681	10 US-09-790-988-1	Sequence 1, Appli
9	18	2.4	18	9 US-10-024-955-3	Sequence 3, Appli
10	18	2.4	18	9 US-10-024-955-5	Sequence 15, Appli
11	18	2.4	26	9 US-10-024-955-8	Sequence 8, Appli
12	18	2.4	251	10 US-09-878-574-11858	Sequence 11858, A
13	18	2.4	263	10 US-09-878-574-5784	Sequence 5784, A
14	18	2.4	302	10 US-09-960-352-690	Sequence 690, App
15	18	2.4	358	10 US-09-764-877-238	Sequence 238, App
16	18	2.4	364	10 US-09-764-877-2500	Sequence 2500, App
17	18	2.4	370	10 US-09-917-800A-964	Sequence 964, App
18	18	2.4	401	9 US-09-946-807-822	Sequence 822, App
19	18	2.4	401	10 US-09-795-668-822	Sequence 822, App

C 20	18	2.4	401	10 US-09-795-668-822	Sequence 822, App
C 21	18	2.4	476	10 US-09-764-877-839	Sequence 839, App
C 22	18	2.4	515	10 US-09-866-562-777	Sequence 777, Appl
C 23	18	2.4	521	10 US-09-764-877-3673	Sequence 3673, App
C 24	18	2.4	521	10 US-09-764-877-3674	Sequence 3674, App
C 25	18	2.4	1366	9 US-09-938-842A-3465	Sequence 3465, Ap
C 26	18	2.4	1497	9 US-09-870-759-79	Sequence 79, Appli
C 27	18	2.4	2600	10 US-09-954-456-1157	Sequence 1157, Ap
C 28	18	2.4	4449	10 US-09-815-242-8386	Sequence 8386, Ap
C 29	18	2.4	8351	10 US-09-917-800A-1199	Sequence 1199, Ap
C 30	18	2.4	32177	10 US-09-764-877-3250	Sequence 3250, Ap
C 31	18	2.4	32207	10 US-09-764-877-3251	Sequence 3, Appli
C 32	18	2.4	34337	10 US-09-741-149-3	Sequence 3, Appli
C 33	18	2.4	46718	10 US-09-816-093-3	Sequence 18, Appli
C 34	18	2.4	63000	10 US-09-780-172-18	Sequence 4, Appli
C 35	18	2.4	172637	10 US-09-805-458A-3	Sequence 1, Appli
C 36	18	2.4	326014	10 US-09-731-231A-3	Sequence 3, Appli
C 37	18	2.4	513509	9 US-09-754-853A-4	Sequence 1, Appli
C 38	18	2.4	684973	10 US-09-263-959-1	Sequence 9556, Ap
C 39	17	2.2	209	10 US-09-878-574-9556	Sequence 249, App
C 40	17	2.2	281	10 US-09-764-860-928	Sequence 928, App
C 41	17	2.2	281	10 US-09-764-860-928	Sequence 381, App
C 42	17	2.2	283	12 US-10-044-090-381	Sequence 5639, Ap
C 43	17	2.2	284	9 US-09-796-692-5639	Sequence 36, Appli
C 44	17	2.2	287	10 US-09-764-878-36	Sequence 170, App
C 45	17	2.2	287	10 US-09-764-860-170	

ALIGNMENTS

RESULT 1
US-10-024-955-6
Sequence 6, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides from
House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
City: Boston
STATE: Massachusetts
COUNTRY: USA
Zip: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 761 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 43..681
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 us-10-024-955-6

Query Match 100.0%; Score 761; DB 9; Length 761;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTTATATCAATAAATCCAAAAACATCTTACAAAAATGAAATTTTGTG 60
 DB 1 GATCTTATATCAATAAATCCAAAAACATCTTACAAAAATGAAATTTTGTG 60
 QY 61 ATTGCTGCCGTGGCATTTTGGCGCTTGGCGATCCAAATTCATGATAAATCACC 120
 DB 61 ATTGCTGCCGTGGCATTTTGGCGCTTGGCGATCCAAATTCATGATAAATCACC 120
 QY 121 GAAGAATCAACAAGCTATTGATGATGCGATGCTGATTTGAACAATCCGAACAATA 180
 DB 121 GAAGAATCAACAAGCTATTGATGATGCGATGCTGATTTGAACAATCCGAACAATA 180
 QY 181 GATCCAAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGATGAT 240
 DB 181 GATCCAAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGATGAT 240
 QY 241 TTCAAGTGAATTAACCATGCGAATTCGAGCTCGAGATTTGAACAAATGAACAG 300
 DB 241 TTCAAGTGAATTAACCATGCGAATTCGAGCTCGAGATTTGAACAAATGAACAG 300
 QY 301 CAAGTGATGCTAATGCTCAAGGTGAAGAGGATTTGTTAAAGCTATTGTTGATCG 360
 DB 301 CAAGTGATGCTAATGCTCAAGGTGAAGAGGATTTGTTAAAGCTATTGTTGATCG 360
 QY 361 GTTCACGATGATATGCTCTCGATGATAATGATTTAGCATACAATTTGGTGATCT 420
 DB 361 GTTCACGATGATATGCTCTCGATGATAATGATTTAGCATACAATTTGGTGATCT 420
 QY 421 CCACGACATGCTCATTTCCGATATTCGAAGATTTTGTGCTTCCCTTGGAAT 480
 DB 421 CCACGACATGCTCATTTCCGATATTCGAAGATTTTGTGCTTCCCTTGGAAT 480
 QY 481 TCTGATGAAGGTAACTAATACATGATCTTTGAAGTACGAAATTCGCTAATGTT 540
 DB 481 TCTGATGAAGGTAACTAATACATGATCTTTGAAGTACGAAATTCGCTAATGTT 540
 QY 541 AACCATATGCTGCTCTTCATCTTGATCAATTTTGGCGTTTATCTGATGAT 600
 DB 541 AACCATATGCTGCTCTTCATCTTGATCAATTTTGGCGTTTATCTGATGAT 600
 QY 601 ACCGATATTTCCAGACACCGTAGCGTAAGGAATGACCAATATTTGGACAGCAT 660
 DB 601 ACCGATATTTCCAGACACCGTAGCGTAAGGAATGACCAATATTTGGACAGCAT 660
 QY 661 AAACGTGAATGGAAAAAATTAACCAATGAGATCATTTTCCAACTGTAACATCT 720
 DB 661 AAACGTGAATGGAAAAAATTAACCAATGAGATCATTTTCCAACTGTAACATCT 720
 QY 721 TTTCTACTGACATAAATAAATTTTATTTTATTTCTCC 761
 DB 721 TTTCTACTGACATAAATAAATTTTATTTTATTTCTCC 761

RESULT 2
 us-10-024-955-1
 ; Sequence 1, Application US/10024955
 ; Patent No. US20020168373A1
 ; GENERAL INFORMATION:
 APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
 TITLE OF INVENTION: Allergenic Proteins and Peptides From
 House Dust Mite and Uses Therefor

; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/024,955
 ; FILING DATE: 19-Dec-2001
 ; CLASSIFICATION: <unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/553,336A
 ; FILING DATE: 10-JUN-1996
 ; APPLICATION NUMBER: US 08/081,540
 ; FILING DATE: 22-JUNE-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jane E. Remillard
 ; REGISTRATION NUMBER: 38,872
 ; REFERENCE/DOCKET NUMBER: IMI-032CP2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)742-4214
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 812 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 68..712
 ; FEATURE:
 ; NAME/KEY: mat.peptide
 ; LOCATION: 119..712
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 us-10-024-955-1
 Query Match 6.2%; Score 47; DB 9; Length 812;
 Best Local Similarity 100.0%; Pred. No. 6.7e-14;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 82 GCCGTTTCGGCTGATCAATTCATGATTAATAATCAGCAAGAAAT 128
 DB 107 GCCGTTTCGGCTGATCAATTCATGATTAATAATCAGCAAGAAAT 153
 RESULT 3
 us-10-024-955-12
 ; Sequence 12, Application US/10024955
 ; Patent No. US20020168373A1
 ; GENERAL INFORMATION:
 APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
 TITLE OF INVENTION: Allergenic Proteins and Peptides From
 House Dust Mite and Uses Therefor
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD, LLP
 STREET: 28 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

```
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/024,955
; FILING DATE: 19-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)742-4214
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-024-955-12
;
; Query Match          3.2%; Score 24; DB 9; Length 24;
; Best Local Similarity 100.0%; Pred. No. 0.019;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 555 TCAATCTGGATCCATTTTGGC 582
DB 1 TCAATCTGGATCCATTTTGGC 24

RESULT 4
US-10-024-955-13/c
; Sequence 13, Application US/10024955
; Patent No. US2002016873A1
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; CORRESPONDENCE ADDRESS: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; ADDRESS: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/024,955
; FILING DATE: 19-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
```

```
;
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-024-955-13
;
; Query Match          2.8%; Score 21; DB 9; Length 28;
; Best Local Similarity 100.0%; Pred. No. 0.61;
; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 664 CGTGAATTCGAAAAAATTAA 684
DB 28 CGTGAATTCGAAAAAATTAA 8

RESULT 5
US-09-880-107-2380
; Sequence 2380, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Schert, Uwe
; TITLE OF INVENTION: Gene Logic, Inc.
; FILE REFERENCE: 44921-5028-MO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2380
; LENGTH: 33206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M64554
US-09-880-107-2380
;
; Query Match          2.8%; Score 21; DB 10; Length 33206;
; Best Local Similarity 100.0%; Pred. No. 0.81;
; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 739 AAAATTTTATTTTATTTCT 759
DB 30390 AAAATTTTATTTTATTTCT 30410

RESULT 6
US-09-938-842A-3199/c
; Sequence 3199, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
```

; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3199
; LENGTH: 1966
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3199

Query Match 2.5%; Score 19; DB 9; Length 1966;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 739 AAAATTTTATTTTATTT 757
|||||
DB 1161 AAAATTTTATTTTATTT 1143

RESULT 7
US-09-754-853A-4
; Sequence 4, Application US/09754853A
; Publication No. US2003005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parsons, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754, 853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174, 880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 4
; LENGTH: 513509
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (111805)..(113968)..(114684)..(115204)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(513509)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 318013_region_A3
US-09-754-853A-4

Query Match 2.5%; Score 19; DB 9; Length 513509;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 742 ATTTTATTTTATTTCTC 760
|||||
DB 93423 ATTTTATTTTATTTCTC 93441

RESULT 8
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: MATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790, 988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160

; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 2.5%; Score 19; DB 10; Length 640681;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTTAT 755
|||||
DB 526284 ATAAATTTTATTTTAT 526266

RESULT 9
US-10-024-955-3
; Sequence 3, Application US/10024955
; Patent No. US20020168373A1
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; HOUSE DUST MITE AND USES THEREFOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHYE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/024, 955
; FILING DATE: 19-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/553, 336A
; FILING DATE: 10-JUN-1996
; APPLICATION NUMBER: US 08/081, 540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-024-955-3

Query Match 2.4%; Score 18; DB 9; Length 18;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GATCCATTCACATGAT 111
|||||
DB 1 GATCCATTCACATGAT 18

RESULT 10
US-10-024-955-15/c
Sequence 15, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-024-955-15
Query Match 2.4%; Score 18; DB 9; Length 18;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GATAAATTCGAACGTCAT 225
|||||
DB 18 GATAAATTCGAACGTCAT 1

RESULT 11
US-10-024-955-8
Sequence 8, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-024-955-8

Query Match 2.4%; Score 18; DB 9; Length 26;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 GATCAATTCACACTGAT 111
|||||
DB 9 GATCAATTCACACTGAT 26

RESULT 12
US-09-878-574-11858/c
Sequence 11858, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO: 11858
LENGTH: 251
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701064991H1
US-09-878-574-11858

Query Match 2.4%; Score 18; DB 10; Length 251;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 110 ATAAATCACCAGAGAA 127
|||||
DB 94 ATAAATCACCAGAGAA 77

RESULT 13

```
US-09-878-574-5784
; Sequence 5784, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 5784
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701097359H1
US-09-878-574-5784
```

```
Query Match          2.4%; Score 18; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      320 AAGGTGAAGAGGTATTG 337
Db      236 AAGGTGAAGAGGTATTG 253
```

```
RESULT 14
; Sequence 690, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 690
; LENGTH: 302
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 04-BOVMS1-008-Q1-E1-A11
US-09-960-352-690
```

```
Query Match          2.4%; Score 18; DB 10; Length 302;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      740 AAATTTTATTTATTT 757
Db      84 AAATTTTATTTATTT 67
```

```
RESULT 15
US-09-764-877-238/C
; Sequence 238, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
```

```
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 238
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (328)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (336)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (342)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-238
```

```
Query Match          2.4%; Score 18; DB 10; Length 358;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      740 AAATTTTATTTATTT 757
Db      134 AAATTTTATTTATTT 117
```

```
RESULT 16
US-09-764-877-2500/C
; Sequence 2500, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2500
; LENGTH: 364
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2500
```

```
Query Match          2.4%; Score 18; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      740 AAATTTTATTTATTT 757
Db      134 AAATTTTATTTATTT 117
```

```
RESULT 17
US-09-917-800A-964/C
; Sequence 964, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-05
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
```



```
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 964
; LENGTH: 370
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1176456
US-09-917-800A-964
```

```
Query Match          2.4%; Score 18; DB 10; Length 370;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 725 ACTGACATATAAATAAA 742
Db 30 ACTGACATATAAATAAA 13
```

```
RESULT 18
US-09-946-807-822/c
; Sequence 822, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 822
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-946-807-822
```

```
Query Match          2.4%; Score 18; DB 9; Length 401;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 740 AAATTTTATTTTATTT 757
Db 269 AAATTTTATTTTATTT 252
```

```
RESULT 19
US-09-795-668-822/c
; Sequence 822, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
```

```
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 822
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-822
```

```
Query Match          2.4%; Score 18; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 740 AAATTTTATTTTATTT 757
Db 269 AAATTTTATTTTATTT 252
```

```
RESULT 20
US-09-795-668-822/c
; Sequence 822, Application US/09795668
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 822
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-822
```

```
Query Match          2.4%; Score 18; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 740 AAATTTTATTTTATTT 757
Db 269 AAATTTTATTTTATTT 252
```

```
RESULT 21
US-09-764-877-839
; Sequence 839, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; PRIOR FILING DATE: 2001-01-17
; PRIOR application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 839
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
```

FEATURE:
NAME/KEY: SITE
LOCATION: (450)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-839

Query Match 2.4%; Score 18; DB 10; Length 476;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTTA 754
Db 234 ATAAATTTTATTTTA 251

RESULT 22
US-09-866-562-77/c
Sequence 77, Application US/09866562
Patent No. US2002009758A1
GENERAL INFORMATION:
APPLICANT: Harlocker, Susan L.
APPLICANT: Bangur, Tongtong
APPLICANT: Klee, Jennifer
APPLICANT: Switzer, Anne
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER.
FILE REFERENCE: 210121.502
CURRENT APPLICATION NUMBER: US/09/866,562
CURRENT FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 96
SEQ ID NO 77
LENGTH: 515
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 395,476
OTHER INFORMATION: n = A,T,C or G
US-09-866-562-77

Query Match 2.4%; Score 18; DB 10; Length 515;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 742 ATTTTATTTTATTTCT 759
Db 61 ATTTTATTTTATTTCT 44

RESULT 23
US-09-764-877-3673
Sequence 3673, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3673
LENGTH: 521
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-3673

Query Match 2.4%; Score 18; DB 10; Length 521;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTTA 754
Db 229 ATAAATTTTATTTTA 246

RESULT 24
US-09-764-877-3674
Sequence 3674, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3674
LENGTH: 521
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-3674

Query Match 2.4%; Score 18; DB 10; Length 521;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTTA 754
Db 229 ATAAATTTTATTTTA 246

RESULT 25
US-09-938-842A-3485/c
Sequence 3485, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 3485
LENGTH: 1366
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-3485

Query Match 2.4%; Score 18; DB 9; Length 1366;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 740 AAATTTTATTTATTT 757
Db 968 AAATTTTATTTATTT 951

RESULT 26
US-09-870-759-79
Sequence 79, Application US/09870759
Patent No. US2002017551A1

```

; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Bacteriophage T270
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (286)..(1374)
; OTHER INFORMATION:
US-09-870-759-79

Query Match                2.4%: Score 18; DB 9; Length 1497;
Best Local Similarity 100.0%: Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      6 TATATCAATACATCA 23
        |||
Db      712 TATATCAATACATCA 729

RESULT 27
US-09-954-456-1157
; Sequence 1157, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1157
; LENGTH: 2600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1157

Query Match                2.4%: Score 18; DB 10; Length 2600;
Best Local Similarity 100.0%: Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      41 AAATGATGAATTTTGT 58

; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Bacteriophage T270
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (286)..(1374)
; OTHER INFORMATION:
US-09-870-759-79

Query Match                2.4%: Score 18; DB 9; Length 1497;
Best Local Similarity 100.0%: Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      6 TATATCAATACATCA 23
        |||
Db      712 TATATCAATACATCA 729

RESULT 28
US-09-815-242-8386/C
; Sequence 8386, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8386
; LENGTH: 4449
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(4449)
US-09-815-242-8386

Query Match                2.4%: Score 18; DB 10; Length 4449;
Best Local Similarity 100.0%: Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      549 TGGTGTCTTTCATCTT 566
        |||
Db      666 TGGTGTCTTTCATCTT 649

RESULT 29
US-09-917-800A-1399
; Sequence 1399, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
```

RESULT 31
 US-09-764-877-3250
 ; Sequence 3250, Application US/09764877
 ; Patent No. US20020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC005
 ; CURRENT APPLICATION NUMBER: US/09/764,877
 ; CURRENT FILING DATE: 2001-01-17

```

RESULT 33
US-09-816-093-3/C
: Sequence 3, Application US/09816093
: Patent No. US20020137180A1
: GENERAL INFORMATION:
: APPLICANT: GAN, weilu et al
: TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: C1001182
: CURRENT APPLICATION NUMBER: US/09/816,093
: CURRENT FILING DATE: 2001-03-26
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 46718
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(46718)
: OTHER INFORMATION: n = A,T,C or G

```

US-09-816-093-3

Query Match 2.4%; Score 18; DB 10; Length 46718;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 742 ATTTTATTTTATTTTCT 759
|||||
DB 18574 ATTTTATTTTATTTTCT 18557

RESULT 34

US-09-780-172-18
; Sequence 18, Application US/09780172
; Patent No. US20020147163A1
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Susan M. Freiler
; APPLICANT: Jacqueline Wylt
; TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-ALPHA EXPRESSION
; FILE REFERENCE: RFS-0159
; CURRENT APPLICATION NUMBER: US/09/780,172
; CURRENT FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 96
; SEQ ID NO 18
; LENGTH: 63000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-780-172-18

Query Match 2.4%; Score 18; DB 10; Length 63000;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 738 TAAATTTTATTTTAT 755
|||||
DB 19556 TAAATTTTATTTTAT 19573

RESULT 35

US-09-805-458A-3/c
; Sequence 3, Application US/09805458A
; Patent No. US20020042100A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS,
; FILE REFERENCE: CLO00722
; CURRENT APPLICATION NUMBER: US/09/805,458A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 172637
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(172637)
; OTHER INFORMATION: n = A,T,C or G
US-09-805-458A-3

Query Match 2.4%; Score 18; DB 10; Length 172637;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 740 AAATTTTATTTTATTT 757
|||||
DB 117607 AAATTTTATTTTATTT 117590

RESULT 36

US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01007
; CURRENT APPLICATION NUMBER: US/09/731,231A
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 326014
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(326014)
; OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

Query Match 2.4%; Score 18; DB 10; Length 326014;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 735 AAATTAATTTTATTTT 752
|||||
DB 226356 AAATTAATTTTATTTT 226339

RESULT 37

US-09-754-853A-4/c
; Sequence 4, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 4
; LENGTH: 513509
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (111805)..(113968),(114684)..(115204)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(513509)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 318013_region_A3
US-09-754-853A-4

Query Match 2.4%; Score 18; DB 9; Length 513509;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 742 ATTTTATTTTATTTCT 759
|||||
DB 35780 ATTTTATTTTATTTCT 35763

RESULT 38
US-09-263-959-1
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
; GENERAL INFORMATION:
; APPLICANT: Hood, Leroy E.
; APPLICANT: Rowen, Lee
; APPLICANT: Koop, Ben F.
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTIL
; NUMBER OF SEQUENCES: 1279
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,959
; FILING DATE: 05-MAR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 920010.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 684973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-263-959-1

Query Match 2.4%; Score 18; DB 10; Length 684973;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 740 AATTTTATTTATTT 757
|||||
Db 472684 AATTTTATTTATTT 472701

RESULT 39
US-09-878-574-9556
; Sequence 9556, Application US/09878574
; Patent No. US2002010548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 9556
; LENGTH: 209
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701102627H1
US-09-878-574-9556

Query Match 2.2%; Score 17; DB 10; Length 209;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 147 TGGCATTGCTGCTATG 163
|||||
Db 34 TGGCATTGCTGCTATG 50

RESULT 40
US-09-764-878-249/C
; Sequence 249, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 249
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-249

Query Match 2.2%; Score 17; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 741 AATTTTATTTATTT 757
|||||
Db 62 AATTTTATTTATTT 46

RESULT 41
US-09-764-860-928/C
; Sequence 928, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 928
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-928

Query Match 2.2%; Score 17; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 741 AATTTTATTTATTT 757
|||||
Db 62 AATTTTATTTATTT 46

RESULT 42
US-10-044-090-381/C
; Sequence 381, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US

;; CURRENT APPLICATION NUMBER: US/10/044,090
;; CURRENT FILING DATE: 2002-01-09
;; NUMBER OF SEQ ID NOS: 850
;; SOFTWARE: PERL Program
;; SEQ ID NO 381
;; LENGTH: 283
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; OTHER INFORMATION: Incyte ID No. US20020137081A1 1379080.1
US-10-044-090-381

Query Match 2.2%: Score 17; DB 12; Length 283;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
|||||
Db 158 AATTTTATTTTATTT 142

RESULT 43
; Sequence 5639, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077,001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5639
; LENGTH: 284
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-5639

Query Match 2.2%: Score 17; DB 9; Length 284;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 741 AATTTTATTTTATTT 757
|||||

Db 123 AATTTTATTTTATTT 107

RESULT 44
US-09-764-878-36/c
; Sequence 36, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 287
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-878-36

Query Match 2.2%: Score 17; DB 10; Length 287;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
|||||
Db 62 AATTTTATTTTATTT 46

RESULT 45
US-09-764-860-170/c
; Sequence 170, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 287
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-860-170

Query Match 2.2%: Score 17; DB 10; Length 287;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
|||||
Db 62 AATTTTATTTTATTT 46

RESULT 46
US-09-796-692-6535/c
; Sequence 6535, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander

```

; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077,001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6535
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-6535

Query Match          2.2%; Score 17; DB 9; Length 318;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 AATTTTATTTTATTT 757
Db 76 AATTTTATTTTATTT 60

RESULT 47
US-09-960-352-4429
; Sequence 4429, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511,006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4429
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 19-LIB34-035-Q1-E1-E11
US-09-960-352-4429
```

```

Query Match          2.2%; Score 17; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 730 CATAAATAAATAATTT 746
Db 346 CATAAATAAATAATTT 362

RESULT 48
US-09-960-352-8048/c
; Sequence 8048, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511,006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8048
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 35-BOVMS1-005-Q1-E1-A4
US-09-960-352-8048

Query Match          2.2%; Score 17; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 668 AATTGGAATAAATAATTA 684
Db 67 AATTGGAATAAATAATTA 51

RESULT 49
US-09-764-853-423
; Sequence 423, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 423
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (41)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (97)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-853-423

Query Match          2.2%; Score 17; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


OY 132 CAAAGCTATTGATGATG 148
|||||
DB 363 CAAAGCTATTGATGATG 379

RESULT 50

US-09-764-898-148
: Sequence 148, Application US/09764898
: Patent No. US20020090673A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PJ201
: CURRENT APPLICATION NUMBER: US/09/764,898
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 311
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 148
: LENGTH: 415
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (17)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (41)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (97)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-764-898-148

Query Match 2.2%; Score 17; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 132 CAAAGCTATTGATGATG 148
|||||
DB 363 CAAAGCTATTGATGATG 379

Search completed: February 21, 2003, 00:28:57
Job time : 1733 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 22:03:38 ; Search time 1482 seconds

(without alignments)
8316.306 Million cell updates/sec

Title: US-10-024-955-6

Perfect score: 761

Sequence: 1 GATCTATATCATATACAAAT.....ATTATTTATTTTATTTCTCC 761

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapept 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hrc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	3.0	570	12	BG131742
2	23	3.0	603	12	BG600512
3	23	3.0	603	13	B0331155
4	23	3.0	635	13	B0331155
5	23	3.0	637	12	BG592911
6	22	2.9	418	13	BM275771

c	7	22	2.9	439	13	BM274902	BM274902
c	8	22	2.9	540	17	DR3N3S	A1748477
c	9	22	2.9	545	17	B0560357	Danio rer
c	10	22	2.9	560	17	A2524638	B0560357
c	11	22	2.9	638	17	A0395435	H4063612-
c	12	21	2.8	192	9	AA180727	A2524638
c	13	21	2.8	374	17	BH290099	C17B1-E1-
c	14	21	2.8	405	12	BG137566	A0395435
c	15	21	2.8	454	12	BF462573	BH290099
c	16	21	2.8	454	12	BH312188	CH230-207
c	17	21	2.8	503	9	AA430887	BH312188
c	18	21	2.8	585	17	BH705338	CH230-101
c	19	21	2.8	591	17	BH711433	AA430887
c	20	21	2.8	684	17	A0319308	RRAMCA148
c	21	21	2.8	701	17	BH428670	BH705338
c	22	21	2.8	726	17	A2997668	BH711433
c	23	21	2.8	730	17	BH462426	BOMLK77TR
c	24	21	2.8	736	17	A0291447	A0319308
c	25	21	2.8	737	17	BH511141	RPCI11-99
c	26	21	2.8	743	17	BH678100	BH428670
c	27	21	2.8	850	17	BH068770	B0GLX14TF
c	28	21	2.8	876	17	A2680222	A2997668
c	29	21	2.8	907	17	BH159038	2M0284118
c	30	21	2.8	1035	13	BM454283	BH462426
c	31	20	2.6	165	13	B1070313	BOHIM01TR
c	32	20	2.6	171	9	AU074661	AQ291447
c	33	20	2.6	222	13	B1125558	BOEG45TF
c	34	20	2.6	240	9	AU071341	BOMAG02TR
c	35	20	2.6	248	13	BM535454	RPCI-24-3
c	36	20	2.6	300	9	AV185077	AZ680222
c	37	20	2.6	300	14	C35776	ENTRK29TR
c	38	20	2.6	324	17	A2772185	BH159038
c	39	20	2.6	344	9	A1356893	BM454283
c	40	20	2.6	344	10	BB579841	B1070313
c	41	20	2.6	391	14	BQ596136	AU074661
c	42	20	2.6	403	17	A2823056	B1125558
c	43	20	2.6	409	9	AU037304	I062892P
c	44	20	2.6	418	9	AU039211	AU071341
c	45	20	2.6	427	14	BQ576836	BM535454

ALIGNMENTS

RESULT 1	BG131742	570 bp	linear	EST 31-JAN-2001
LOCUS	EST464634	tomato crown gall	Lycopersicon	esculentum
DEFINITION	EST464634	5' sequence, mRNA	sequence.	
ACCESSION	BG131742			
VERSION	BG131742.1	GI:12631930		
KEYWORDS	EST.			
SOURCE	tomato.			
ORGANISM	Lycopersicon	esculentum		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;			
	Lycopersicon.			
	1 (bases 1 to 570)			
	van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning			
	,C. and Tanksley,S.			
	Generation of ESTs from tomato crown gall tissue			
TITLE	Unpublished (2001)			
JOURNAL	Contact: CUGI			
COMMENT	Clemson University Genomics Institute			
	Clemson University			
	100 Jordan Hall, Clemson, SC 29634, USA			
	Email: http://www.genome.clemson.edu/orders/index.html .			
FEATURES	Location/Dualifiers			
source	1..570			
	/organism="Lycopersicon esculentum"			
	/cultivar="TA496"			
	/db_xref="taxon:4081"			

```

/clone="cTOE5121"
/clone_lib="tomato crown gall"
/tissue_type="crown gall"
/dev_stage="crown galls from full-grown plants (8 wks old)"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Four wk old greenhouse plants were stab inoculated
on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
Cornell U.). Galls were allowed to develop for another 4
wks, when gall tissue was frozen in liquid nitrogen."
BASE COUNT      159 a      104 c      145 g      162 t
ORIGIN
Query Match      3.0%; Score 23; DB 12; Length 570;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      629 AGGAATGACCAAGTATTGCA 651
        |||||||
Db      437 AGGAATGACCAAGTATTGCA 459

RESULT 2
BG600512      603 bp      mRNA      linear      EST 12-APR-2001
LOCUS
DEFINITION
EST505407 csts Solanum tuberosum cDNA clone CTS29C2 5' sequence,
mRNA sequence.
ACCESSION
VERSION      BG600512.1      GI:13617648
KEYWORDS
SOURCE
ORGANISM
Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE
AUTHORS
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
Bougril,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
Generations of ESTs from sprouting potato eyes
JOURNAL
Unpublished (2000)
COMMENT
Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES
Source
Location/Qualifiers
1..603
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CTS29C2"
/clone_lib="csts"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."
BASE COUNT      162 a      112 c      160 g      169 t
ORIGIN
Query Match      3.0%; Score 23; DB 12; Length 603;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      629 AGGAATGACCAAGTATTGCA 651
        |||||||
Db      477 AGGAATGACCAAGTATTGCA 499

```

```

RESULT 3
BJ331155      603 bp      mRNA      linear      EST 05-MAR-2002
LOCUS
DEFINITION
BJ331155 Dictyostellium discoideum cDNA library, AF Dictyostellium
discoideum cDNA clone dda34102 5', mRNA sequence.
ACCESSION
VERSION      BJ331155.1      GI:19161285
KEYWORDS
SOURCE
Dictyostellium discoideum.
ORGANISM
Dictyostellium discoideum.
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
REFERENCE
AUTHORS
1 (bases 1 to 603)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shih-I,T.
Full length cDNA of Dictyostellium discoideum at the aggregation
stage
JOURNAL
Unpublished (2002)
Contact: Tadasu Shih-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES
Source
Location/Qualifiers
1..603
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda34102"
/clone_lib="Dictyostellium discoideum cDNA library, AF"
/sex="mat A"
/dev_stage="Aggregation stage"
BASE COUNT      215 a      86 c      74 g      225 t      3 others
ORIGIN
Query Match      3.0%; Score 23; DB 13; Length 603;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      735 AATTAATTTTATTATTATT 757
        |||||||
Db      109 AATTAATTTTATTATTATT 131

RESULT 4
BM110995      635 bp      mRNA      linear      EST 26-NOV-2001
LOCUS
DEFINITION
EST538531 potato roots Solanum tuberosum cDNA clone cPRO9L22 5' end
, mRNA sequence.
ACCESSION
VERSION      BM110995.1      GI:17073175
KEYWORDS
SOURCE
EST.
ORGANISM
Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE
AUTHORS
1 (bases 1 to 635)
van der Hoeven,R., Sun,H., Karmycheva,S.A., Tsai,J., Van Aken,S.,
Utterback,T., Chiemiango,A., Bougril,O., Buell,C.R., Ronning,C.,
Tanksley,S. and Baker,B.
Generation of ESTs from potato roots
JOURNAL
Unpublished (2001)
Contact: Research Genetics, Libraries Division
Tel: 1-800-711-6195
Email: cdna@resgen.com
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: T3

FEATURES
Source
Location/Qualifiers
1..635
/organism="Solanum tuberosum"

```

```

/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cPRO9L22"
/clone_lib="potato roots"
/tissue_type="roots"
/dev_stage="in vitro grown stem cuttings"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University, Tanksley lab;
sequencing: The Institute for Genomic Research. Roots were
isolated from in vitro grown stem cuttings on CM medium.
Roots were isolated two weeks after placing the stem
cuttings from in vitro grown plants on medium."
BASE COUNT      168 a      126 c      165 g      176 t
ORIGIN
Query Match      3.0%; Score 23; DB 13; Length 635;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 AGGAATGACCAAGTATTGCA 651
|||||
Db 469 AGGAATGACCAAGTATTGCA 491

RESULT 5
LOCUS      BG592911      637 bp      mRNA      linear      EST 12-APR-2001
DEFINITION EST491589 cSTS Solanum tuberosum cDNA clone cSTS2D10 5' sequence,
mRNA sequence.
ACCESSION  BG592911
VERSION     BG592911.1 GI:13611051
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
            1 (bases 1 to 637)
REFERENCE   1 (bases 1 to 637)
AUTHORS    van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
            Generations of ESTs from sprouting potato eyes
            Unpublished (2000)
COMMENT     Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdaeresgen.com
            Seq primer: M13F-R.
FEATURES
    source
        Location/Qualifiers
            1..637
                /organism="Solanum tuberosum"
                /cultivar="Kennebec"
                /db_xref="taxon:4113"
                /clone="cSTS2D10"
                /clone_lib="cSTS"
                /tissue_type="sprouting eyes from tubers"
                /dev_stage="12-14 weeks post harvest"
                /lab_host="SOLR"
            /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
            XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
            taken from tubers. The tubers were incubated at 26C in the
            dark for 2-3 weeks prior to sprouting. The eyes were
            frozen in liquid nitrogen immediately upon removal from
            tubers."
BASE COUNT      164 a      125 c      164 g      184 t
ORIGIN
Query Match      3.0%; Score 23; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 AGGAATGACCAAGTATTGCA 651
|||||

```

```

Db 486 AGGAATGACCAAGTATTGCA 508

RESULT 6
LOCUS      BM275771/c      418 bp      mRNA      linear      EST 20-DEC-2001
DEFINITION pFESToa85d08.y1 Plasmodium falciiparum 3D7 gametocyte cDNA library
Plasmodium falciiparum 3D7 cDNA 5', mRNA sequence.
ACCESSION  BM275771
VERSION     BM275771.1 GI:17969111
KEYWORDS   EST.
SOURCE     Plasmodium falciiparum 3D7.
            Plasmodium falciiparum 3D7.
ORGANISM   Plasmodium falciiparum 3D7.
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
            1 (bases 1 to 418)
REFERENCE   1 (bases 1 to 418)
AUTHORS    Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
            Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
            Bowers,Y., Gibbons,M., Rifter,E., Bennett,J., Jentes,E., Ronko,I.,
            Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
            Maguire,L., Richey,J., Madkins,J., Kennedy,S., Levinso,D.,
            Waterston,R., Wilson,R. and Sibley,D.
            Washu Plasmodium EST Project
            Unpublished (2001)
COMMENT     Contact: L. David Sibley
            Washu Plasmodium EST Project
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watsen.wustl.edu
            Library was constructed by R. Haywood. DNA sequencing by:
            Washington University Genome Sequencing Center for information on
            obtaining a clone please contact: L. David Sibley
            (sibley@borcim.wustl.edu), Washington University
            Seq primer: -400P from Glibco
            High quality sequence stop: 385.
FEATURES
    source
        Location/Qualifiers
            1..418
                /organism="Plasmodium falciiparum 3D7"
                /db_xref="taxon:36329"
                /clone_lib="Plasmodium falciiparum 3D7 gametocyte cDNA
                library"
                /dev_stage="gametocyte (stage III-V)"
                /lab_host="DH10B (Genesig, Invitrogen, Inc.)"
            /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
            XhoI; The library was constructed by R Haywood. cDNAs were
            synthesized from gametocyte poly(A)+ RNA by oligo d(T)
            priming, size-selected and directionally cloned into the
            EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
            lambda vector (Stratagene). The primary library was mass
            excised as phagemid using the Exassist helper phage
            (Stratagene). Clones were mass excised using the Exassist
            helper phage (Stratagene), the phagemids were precliplated
            with PEG 8000 and extracted with phenol/chloroform.
            Phagemid DNA was electroporated into DH10B cells. Clone
            Availability: David Sibley, Washington University."
BASE COUNT      222 a      26 c      8 g      161 t
ORIGIN
Query Match      2.9%; Score 22; DB 13; Length 418;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 ATAAATATAATTTTATTATT 753
|||||
Db 313 ATAAATATAATTTTATTATT 292

RESULT 7
LOCUS      BM274902/c      439 bp      mRNA      linear      EST 20-DEC-2001
DEFINITION pFESToa876f12.y1 Plasmodium falciiparum 3D7 gametocyte cDNA library
Plasmodium falciiparum 3D7 cDNA 5', mRNA sequence.

```

ACCESSION BM274902 GI:17968220
 VERSION BM274902.1
 KEYWORDS EST.
 SOURCE Plasmodium falciparum 3D7.
 ORGANISM Plasmodium falciparum 3D7.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 439)
 TANG, K., COLE, R., CHAKRABARTI, D., HAYWOOD, R., CLIFTON, S., PAPE, D.,
 MARA, M., HILLIER, L., MARTIN, J., WYLLIE, T., DANTE, M., THEISING, B.,
 BOWERS, Y., GIBBONS, M., RITTER, E., BENNETT, J., JENTES, E., RONKO, I.,
 TSAGARELISHVILI, R., BELAYGROD, L., FRANKLIN, C., CARR, L., GROW, A.,
 MAGUIRE, L., RICHEY, J., MADKINS, J., KENNEDY, S., LEVINSO, D.,
 WATERSTON, R., WILSON, R. and SIBLEY, D.
 TITLE Washu Plasmodium EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: L. David Sibley
 Washu Plasmodium EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Library was constructed by R. Haywood. DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: L. David Sibley
 (sibley@borcim.wustl.edu), Washington University
 Seq primer: -400P from Gibco
 High quality sequence stop: 363.
 FEATURES
 source
 location/Qualifiers
 1..439
 /organism="Plasmodium falciparum 3D7"
 /db_xref="taxon:36329"
 /clone_lib="Plasmodium falciparum 3D7 gametocyte CDNA
 library"
 /dev_stage="gametocyte (stage III-V)"
 /lab_host="DH10B (Genesig, Invitrogen, Inc.)"
 /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
 XhoI; The library was constructed by R Haywood. cDNAs were
 synthesized from gametocyte poly(A)+ RNA by oligo d(T)
 priming, size-selected and directionally cloned into the
 EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
 lambda vector (Stratagene). The primary library was mass
 excised as phagemid using the Exsist helper phage
 (Stratagene). Clones were mass excised using the Exsist
 helper phage (Stratagene), the phagemids were precipitated
 with PBG 8000 and extracted with phenol/chloroform.
 Phagemid DNA was electroporated into DH10B cells. Clone
 Availability: David Sibley, Washington University."
 BASE COUNT 232 a 31 c 162 t
 ORIGIN
 Query Match 2.9%; Score 22; DB 13; Length 439;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 732 ATAAATTAATTTTATTTT 753
 ||||||||||||||||||
 Db 313 ATAAATTAATTTTATTTT 292
 RESULT 8
 DR3N3S/C 540 bp DNA linear GSS 06-JUN-2002
 LOCUS Danio rerio genomic clone DKey-3N3, genomic survey sequence.
 ACCESSION AL748477
 VERSION AL748477.1 GI:21350358
 KEYWORDS GSS.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 540)

AUTHORS Humphray, S.J., Huckle, E. and Hunt, S.E.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
 humphreys@sanger.ac.uk Unpublished
 COMMENT This sequence was generated from the SP6 end of BAC 3N3. 3N3 is
 part of the Daniokey Pilot BAC Library created by R. Plasterk and
 N.V. Keygene.
 Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
 FEATURES
 source
 location/Qualifiers
 1..540
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="DKey-3N3"
 /issue_type="Testis"
 /note="Vector pIndigoBAC-536"
 BASE COUNT 227 a 44 c 45 g 224 t
 ORIGIN
 Query Match 2.9%; Score 22; DB 17; Length 540;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 733 TAAATTAATTTTATTTT 754
 ||||||||||||||||||
 Db 172 TAAATTAATTTTATTTT 151
 RESULT 9
 B0560357/c 545 bp mRNA linear EST 20-JUN-2002
 LOCUS H4063G12-3 NIA Mouse 7.4K CDNA Clone Set Mus musculus CDNA clone
 DEFINITION H4063G12 3', mRNA sequence.
 ACCESSION B0560357
 VERSION B0560357.1 GI:21461242
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 545)
 VANBUREN, V., PIAO, Y., DUDKULA, D.B., QIAN, Y., CARTER, M.G., MARTIN,
 P.R., STAGG, C.A., BASSEY, D., ALBA, K., HAMATANI, T., KARGUL, G.J.,
 LUO, A.G. and KO, M.S.H.
 Assembly, verification, and initial annotation of NIA 7.4K mouse
 CDNA clone set
 JOURNAL Unpublished (2002)
 COMMENT Other_ESTS: H4063G12-5
 Contact: Yong Qian
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit http://lgsun.grc.nia.nih.gov/cDNA/NIA_7.4k.html for details.
 Plate: H4063 row: G column: 12
 Seq primer: -21M13 Forward
 High quality sequence stop: 545
 POLY-A=yes.
 FEATURES
 source
 location/Qualifiers
 1..545
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="nia:EST:H4063G12-3"
 /db_xref="taxon:10090"
 /clone_lib="H4063G12"
 /clone_lib="NIA Mouse 7.4K CDNA Clone Set"
 /sex="mixed"
 /dev_stage="mixed"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
 clone is among a rearranged set of 7,407 clones from more
 than 20 CDNA libraries."

BASE COUNT 198 a 94 c 82 g 171 t
ORIGIN

Query Match 2.9%: Score 22; DB 14; Length 545;
Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 661 AATCGTAAATGAAAAAATT 682
|||||
Db 207 AACGTGAATGAAAAAATT 186

RESULT 10
A2524638 560 bp DNA linear GSS 07-MAY-2001
LOCUS

DEFINITION 234bpb09 Pb MBN #21 Plasmodium berghei genomic 3', DNA sequence.

ACCESSION A2524638
VERSION A2524638.1 GI:13964714

KEYWORDS GSS.

SOURCE Plasmodium berghei.

ORGANISM Plasmodium berghei.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 560)

AUTHORS Carlton, J.M., R. and Dame, J.B.

TITLE The Plasmodium vivax and P. berghei gene sequence tag projects

JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)

COMMENT

CONTACT: Dame JB

Dept. of Pathobiology, College of Veterinary Medicine

University of Florida

2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA

Tel: 352 392 4700

Fax: 352 392 9704

Email: damej@mail.ufl.edu

Seq primer: M13(-20) forward

Class: Shotgun.

FEATURES

Location/Qualifiers

1..560

/organism="Plasmodium berghei"

/strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"

/db_xref="taxon:5821"

/clone_1lib="Pb MBN #21"

/dev_stage="asexual blood forms"

/lab_host="Mus musculus"

/note="Vector: pBluescript SK(+) vector DNA, phagemid

excised from lambda ZAP; Site_1: EcoRV; Site_2: EcoRV;

Genomic DNA was prepared from asynchronous blood stage

forms of the cloned ANKA isolate of P. berghei grown in

laboratory Swiss white mice. The DNA was purified from

contaminating host DNA by Hoechst Dye 33258-CsCl

ultracentrifugation and precipitated. Purified DNA was

digested with mung bean nuclease in the presence of 36-38%

formamide at 50 C, as described (Vernick, K.D., Imberski,

R.B., and McCutchan, T.F. 1988. Nucleic Acids Research

16:6883-6896). The ends of the digestion fragments were

polished using T4 DNA polymerase, and the fragments size

selected in the range 500-2000 bp. These were ligated into

the EcoRV-cleaved and dephosphorylated pBluescript SK(+) vector.

Recombinant plasmids were used to transform E. coli XL10-Gold host cells."

BASE COUNT 212 a 44 c 63 g 241 t

ORIGIN

Query Match 2.9%: Score 22; DB 17; Length 560;
Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 731 AATAAAATTAATTTTATTTT 752
|||||
Db 345 AATAAAATTAATTTTATTTT 366

RESULT 11

AQ395435/c

LOCUS AQ395435 638 bp DNA linear GSS 06-MAR-1999
DEFINITION CITR1-El-2546N18.TR CITR1-El Homo sapiens genomic clone 2546N18,
DNA sequence.

ACCESSION AQ395435

VERSION AQ395435.1 GI:4366461

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 638)

Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and

Venter, J.C.

Use of BAC End Sequences from Caltech Libraries for Sequence-Ready

Map Building

Unpublished (1997)

CONTACT: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tldb/hungun/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES

Location/Qualifiers

1..638

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="2546N18"

/clone_1lib="CITR1-El"

/sex="male"

/cell_type="sperm"

/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;

Caltech Human BAC Library D"

BASE COUNT 253 a 116 c 110 g 159 t

ORIGIN

Query Match 2.9%: Score 22; DB 17; Length 638;
Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 731 AATAAAATTAATTTTATTTT 752
|||||
Db 219 AATAAAATTAATTTTATTTT 198

RESULT 12

AA180727

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

EST.

Brugia malayi.

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;

Onchocercidae; Brugia.

1 (bases 1 to 192)

Blaxter, M.L., Waterfall, M., Daub, J., Lizotte, M., Baron, L. and Jones

, S.J.

Genes expressed in adult female Brugia malayi

Unpublished (1996)

CONTACT: Blaxter ML

Institute of Cell, Animal and Population Biology

University of Edinburgh

Asmworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9

3JF, UK.

Tel: +44 131 650 6760

Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 The ABI trace of this sequence can be viewed at
<http://www.sanger.ac.uk/Druga/AF/MBAPG3H06G3.html> This is the
 full sequence of the cDNA clone. The polyA tail has been clipped
 and is excluded from this sequence
 Seq primer: T3.

FEATURES

source

Location/Qualifiers
 1..192
 /organism="Brugia malayi"
 /db_xref="taxon:6279"
 /clone="AFG3H06"
 /clone_1lb="Brugia malayi adult female cDNA (SAM96MLM-BmaF)"
 /sex="female"
 /dev_stage="adult"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
 Xho I; Lymphatic filarial nematode parasite of humans.
 mRNA was prepared from approximately 50 adult females
 isolated from the peritoneal cavity of jirds and
 converted to double-stranded cDNA using reverse
 transcriptase and oligo(dT) followed by RNase H and DNA
 pol I. The library has 5 x 10⁶ independent recombinants
 and the average insert size is ~900bp. The library was
 constructed by Michelle Lizotte-Maniowski. The
 library is available from Dr.S.A.Williams, email:
 genome@smith.edu."

BASE COUNT 93 a 23 c 32 g 44 t
 ORIGIN

Query Match 2.8%; Score 21; DB 9; Length 192;
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 668 AATTGGAATAAATTACCAA 688

Db 148 AATTGGAATAAATTACCAA 168

RESULT 13

BH290099

LOCUS

BH290099 374 bp DNA linear GSS 30-NOV-2001
 CH230-207L5.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone

ACCESSION

BH290099

BH290099.1 GI:17202507

KEYWORDS

SOURCE

ORGANISM

Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 374)
 Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
 'A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
 Jong,P. and Fraser,C.M.
 Rat BAC End Sequences from Library CHORI-230 Ecort segment
 Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
<http://www.chori.org/bacpac/rat230.htm>. For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
http://www.chori.org/bacpac/orering_information.htm. BAC end
 page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html
 Plate: 207 row: L column: 5

Seq primer: SP6
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..374
 /organism="Rattus norvegicus"
 /strain="BN/SSNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-207L5"
 /clone_1lb="CHORI-230 Segment 1"
 /sex="Female"
 /cell_type="Brain"
 /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
 CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
 Pieter de Jong"

BASE COUNT 114 a 68 c 61 g 131 t
 ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 374;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 AAAAAACATCTTACAAA 43

Db 38 AAAAAACATCTTACAAA 58

RESULT 14

BG137566

LOCUS

BG137566 405 bp mRNA linear EST 31-JAN-2001
 EST478008 wild tomato pollen Lycopersicon pennellii cDNA clone

ACCESSION

BG137566

VERSION

KEYWORDS

SOURCE

ORGANISM

Lycopersicon pennellii.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

REFERENCE

AUTHORS

1 (bases 1 to 405)
 van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Uterback,T.,
 Hansen,C., Romling,C. and Tanksley,S.
 Generation of ESTs from wild tomato (L. pennellii) pollen
 Unpublished (2001)

COMMENT

CONTACT

Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>.
 Location/Qualifiers

FEATURES

source

1..405
 /organism="Lycopersicon pennellii"
 /cultivar="TA56"
 /db_xref="taxon:28526"
 /clone="CLP6F5"
 /clone_1lb="wild tomato pollen"
 /tissue_type="pollen"
 /dev_stage="pollen collected from open flowers"
 /lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; Pollen was collected from open flowers from
 L.pennellii TA56, and stored at -80 C until library
 construction."

TITLE

JOURNAL

COMMENT

BASE COUNT 121 a 98 c 67 g 119 t
 ORIGIN

Query Match 2.8%; Score 21; DB 12; Length 405;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 276 TCGAGATTGAAACAATGAA 296

|||||

Db 151 TCGAGGATTGAACAATGAA 171

RESULT 15
LOCUS BF462573
DEFINITION UI-M-CCOP-bnc-b-06-0-UI.s1 NIH_BMAP_Ret4.S2 Mus musculus cDNA clone
ACCESSION BF462573
VERSION BF462573.1 GI:11531756
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 411)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mest@mail.nih.gov
Oligo-ot track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENERICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
The following repetitive elements were found in this cDNA sequence:
22-87, >POLY_A#Simple_repeat
Seq primer: M13 Forward
POLY-A-No.

FEATURES
source location/Qualifiers
1..411
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UI-M-CCOP-bnc-b-06-0-UI"
/lab_host="NIH_BMAP_Ret4.S2"
/note="Vector: p17T3D-Pac (Pharmacia)"
polylinker: Site.1: Not I; Site.2: Eco RI; The
NIH_BMAP_Ret4.S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu. The tissue
for this library was contributed by Dr. Xin-Yuan Fu, Yale
University School of Medicine
TAG_SEQ=None found"

BASE COUNT 129 a 82 c 79 g 121 t
ORIGIN

Query Match 2.8% Score 21; DB 12; Length 411;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 668 AATGGAAAAAATTAACCA 688
|||||
Db 148 AATTGGAAAAAATTAACCA 168

RESULT 16
LOCUS BH312188
DEFINITION CH230-101J16.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone

CH230-101J16, DNA sequence.
BH312188
GI:17239932

ACCESSION
VERSION BH312188.1
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 454)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcORI segment
COMMENT Unpublished (1999)
Other_GSSs: CH230-101J16.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/orering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac-ends/rat/bac_end_intro.html
Plate: 101 row: J column: 16
Seq primer: SP6
Class: BAC ends.

FEATURES
source location/Qualifiers
1..454
/organism="Rattus norvegicus"
/strain="BN/SSNhsd/MCW"
/db_xref="taxon:10116"
/clone_lib="CH230-101J16"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site.1: EcoRI; Site.2: EcoRI;
CHORI-230 Rat (BN/SSNhsd/MCW) BAC library produced by
Pieter de Jong"

BASE COUNT 114 a 91 c 84 g 165 t
ORIGIN

Query Match 2.8% Score 21; DB 17; Length 454;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 ATAAATAAATTTTATTTT 752
|||||
Db 238 ATAAATAAATTTTATTTT 218

RESULT 17
LOCUS AA430887
DEFINITION RRAMCA1483SK Brugia malayi adult male cDNA (SAM94NL-Bmaw) Brugia
malayi cDNA clone RRAMCA1483 5', mRNA sequence.
ACCESSION AA430887
VERSION AA430887.1 GI:2113569
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.

REFERENCE 1 (bases 1 to 503)
AUTHORS Williams,S.A.
TITLE Genes expressed in adult males of Brugia malayi
JOURNAL Unpublished (1995)
COMMENT Contact: Steven A. Williams

Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seg primer: bluescript SK.

FEATURES
source
1. 503
Location/Qualifiers

/organism="Brugia malayi"
/strain="FRS Labs"
/db_xref="taxon:6279"
/clone_lib="Brugia malayi adult male cDNA (SAW94NL-BMAM)"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda UniTap XR, Site_1: Ecor I; Site_2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from adult males of Brugia malayi
isolated from birds and converted to double stranded cDNA
using reverse transcriptase and oligo(dT) followed by
RNase H and DNase I. The library had 4.6 x 10⁶
independent recombinants and average insert size was 800
base pairs. The library was constructed by Noelle Ling.
The library is available from Dr. S.A. Williams, email
genome@smith.edu."

BASE COUNT 168 a 74 c 95 g 165 t 1 others
ORIGIN

Query Match 2.8%; Score 21; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 42 ATGATGCAATTTTGTGTGAT 62
|||||
Db 207 AATGATCAAAATTTTGTGTGAT 227

RESULT 18 585 bp DNA linear GSS 20-FEB-2002
BH705338 BOHWM54TF BO_2_3_KB Brassica oleracea genomic clone BOHWM54, DNA
LOCUS
DEFINITION
sequence.
ACCESSION BH705338
VERSION BH705338.1 GI:18787883
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM

Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 585)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
Other_GSSs: BOHWM54TR
COMMENT
Contact: Chris Town

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
1. 585
Location/Qualifiers

/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOHWM54"
/clone_lib="BO_2_3_KB"
/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared

BASE COUNT 247 a 60 c 45 g 233 t
ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 585;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 737 ATAAATTTTATTTTATTT 757
|||||
Db 197 ATAAATTTTATTTTATTT 217

RESULT 19 591 bp DNA linear GSS 20-FEB-2002
BH711433 BOHMK777R BO_2_3_KB Brassica oleracea genomic clone BOHMK77, DNA
LOCUS
DEFINITION
sequence.
ACCESSION BH711433
VERSION BH711433.1 GI:18801445
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM

Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 591)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
Other_GSSs: BOHMK77TF
COMMENT
Contact: Chris Town

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
1. 591
Location/Qualifiers

/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOHMK77"
/clone_lib="BO_2_3_KB"
/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 252 a 65 c 42 g 232 t
ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 591;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 737 ATAAATTTTATTTTATTT 757
|||||
Db 289 ATAAATTTTATTTTATTT 309

RESULT 20 684 bp DNA linear GSS 06-MAY-1999
AQ319308 APC111-99N14.TV RPCI-11 Homo sapiens genomic clone RPCI-11-99N14,
LOCUS
DEFINITION
DNA sequence.
ACCESSION AQ319308
VERSION AQ319308.1 GI:4052273
KEYWORDS GSS.
SOURCE human.
ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 684)
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Mible,C., de Jong,P. and Venter,J.C.
 TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
 JOURNAL Unpublished (1998)
 COMMENT Other GSSs: RPC111-99N14.TJ
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0208
 Fax: 301 838 0208
 Email: hbeetigr.org

Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: T7
 Class: BAC ends.

FEATURES

SOURCE

Location/Qualifiers
 1..684
 /organism="Homo sapiens"
 /db_xref="GDB:7537957"
 /db_xref="taxon:9606"
 /clone="RPC1-11-99N14"
 /clone_lib="RPC1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC6.6; Site_1: EcoRI; Site_2: EcoRI; Rpci11 Human Male BAC library"

BASE COUNT 225 a 94 c 110 g 255 t

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 684;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 739 AAAATTTTATTTATTTCT 759
 ||||||||||||||||||||
 Db 426 AAAATTTTATTTATTTCT 446

RESULT 21

BH428670/c

LOCUS BH428670 701 bp DNA linear GSS 12-DEC-2001
 DEFINITION BOGLX14TF BOGL Brassica oleracea genomic clone BOGLX14, DNA sequence.

ACCESSION BH428670

VERSION BH428670.1

KEYWORDS GI:17614391

SOURCE GSS.

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 701)

AUTHORS Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other GSSs: BOGLX14TR

CONTACT: Chris Town

TIGR Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers
 1..701

/organism="Brassica oleracea"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOGLX14"
 /clone_lib="BOGL"
 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 278 a 63 c 79 g 281 t

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 701;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 737 ATAAATTTTATTTTATTTT 757
 ||||||||||||||||||||

Db 522 ATAAATTTTATTTTATTTT 502

RESULT 22

AZ997668/c

LOCUS AZ997668 726 bp DNA linear GSS 27-APR-2001
 DEFINITION 2M0284118F Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0284118 F, DNA sequence.

ACCESSION AZ997668

VERSION AZ997668.1

KEYWORDS GI:13868895

SOURCE GSS.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 726)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

CONTACT: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0284 row: 1 column: 18

Seq primer: CGTTGTAAACGACGCCACG

Class: plasmid ends

High quality sequence stop: 726.

Location/Qualifiers

1..726

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0284118"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv. Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

electrophoresis. Vector DNA was prepared from a derivative

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 219 a 132 c 143 g 232 t
ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 726;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 723 TCACGTGACATATAATTAAT 743
|||||
DB 282 TCACGTGACATATAATTAAT 262

RESULT 23

BH462426 730 bp DNA linear GSS 13-DEC-2001
LOCUS BOHIM01TR BOH1 Brassica oleracea genomic clone BOHIM01, DNA
DEFINITION

ACCESSION BH462426.1 GI:17654213

VERSION BH462426
KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 730)

REFERENCE Tom, C.D., Van Aken, S., Uteerback, T. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)

COMMENT Other_GSSs: BOHIM01TR
Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR

Class: sheared ends.

FEATURES Location/Qualifiers

1..730
/organism="Brassica oleracea"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOHIM01"

/note="Vector: PHOS1, Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 233 a 116 c 111 g 270 t

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 730;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 737 ATAAATTTTATTTTATTTT 757
|||||

DB 499 ATAAATTTTATTTTATTTT 519

RESULT 24

AQ291447 736 bp DNA linear GSS 03-DEC-1998
LOCUS AQ291447/c
DEFINITION nbx00023J12r CUGI Rice BAC library Oryza sativa genomic clone
KEYWORDS GSS.
ACCESSION AQ291447

ORGANISM

VERSION AQ291447.1 GI:3948705
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 736)

REFERENCE Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293

Email: rwing@clemson.edu
Seq primer: GGAAACGCTATGACCATG
Class: BAC ends
High quality sequence start: 72
High quality sequence stop: 352.

FEATURES Location/Qualifiers

1..736
/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db_xref="taxon:4530"

/clone="nbx0023J12r"

/clone_11b="CUGI Rice BAC library"

/tissue_type="leaf"

/lab_host="E. coli DH10B"

/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 270 a 153 c 155 g 157 t

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 736;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 ATGAATTTTGTGATGCT 66
|||||

DB 577 ATGAATTTTGTGATGCT 557

RESULT 25

BH511141 737 bp DNA linear GSS 13-DEC-2001
LOCUS BH511141/c
DEFINITION BOGE Brassica oleracea genomic clone BOGEG45, DNA
ACCESSION BH511141
VERSION BH511141.1 GI:17719231
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 737)

REFERENCE
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOGEG45FR
COMMENT
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
Location/Qualifiers
1..737
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGEG45"
/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT
289 a 147 c 95 g 206 t

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 737;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 AATAAATTAATTTTATT 751
|||||
Db 517 AATAAATTAATTTTATT 497

RESULT 26
BH678100/c 743 bp DNA linear GSS 19-FEB-2002
LOCUS
BOMAG02R BO_2_3_KB Brassica oleracea genomic clone BOMAG02, DNA
sequence.
ACCESSION
BH678100
VERSION
BH678100.1 GI:18748543
KEYWORDS
GSS.
SOURCE
Brassica oleracea.
ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 743)

REFERENCE
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
Location/Qualifiers
1..743
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOMAG02"
/clone_id="BO_2_3_KB"
/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 301 a 58 c 84 g 300 t

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 743;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 737 AATAAATTTTATTATT 757
|||||
Db 317 AATAAATTTTATTATT 297

RESULT 27
BH068770/c 850 bp DNA linear GSS 18-JUL-2001
LOCUS
RPC1-24-361J7.TJ RPC1-24 Mus musculus genomic clone RPC1-24-361J7,
DNA sequence.
ACCESSION
BH068770
VERSION
BH068770.1 GI:14888367
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 850)

REFERENCE
ZHAO,S., NIERMAN,W., MALEK,J., SHATSMAN,S., AKINREI,B., LEVINS,M., TSEGAYE,G., GEER,K., KROL,M., SHVARTSBEYN,A., GEORGEORGIS,E., RUSSELL,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPC1-24
Unpublished (1999)
Other GSSs: RPC1-24-361J7.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPC1-24. For BAC library availability, please contact Pieter de Jong (pdejong@tigr.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 361 row: J column: 7
Seq primer: SP6
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..850
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-24-361J7"
/clone_id="RPC1-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI; RPC1-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pPARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 286 a 114 c 132 g 318 t

ORIGIN

Query Match 2.8%; Score 21; DB 17; Length 850;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 AATAAATTAATTTTATT 751
|||||
Db 577 AATAAATTAATTTTATT 557

RESULT 28

AZ680222 876 bp DNA linear GSS 14-DEC-2000
LOCUS ENTAK29TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
COMMENT genomic, DNA sequence.
ACCESSION AZ680222
VERSION AZ680222.1 GI:11817368
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 876)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igf.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 16
High quality sequence stop: 790.
FEATURES
source
1. 876
Location/Qualifiers
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS01, Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999)."

BASE COUNT 379 a 83 c 72 g 342 t
ORIGIN
Query Match 2.8%; Score 21; DB 17; Length 876;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 731 AATAAATAAATTTTATT 751
|||||
Db 19 AATAAATAAATTTTATT 39

RESULT 29
LOCUS BH159038 907 bp DNA linear GSS 24-SEP-2001
DEFINITION ENTST50TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
COMMENT genomic, DNA sequence.
ACCESSION BH159038
VERSION BH159038.1 GI:15732476
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 907)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igf.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 7
High quality sequence stop: 635.
FEATURES
source
1. 907
Location/Qualifiers
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS01, Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999)."

BASE COUNT 341 a 117 c 70 g 379 t
ORIGIN
Query Match 2.8%; Score 21; DB 17; Length 907;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 731 AATAAATAAATTTTATT 751
|||||
Db 287 AATAAATAAATTTTATT 267

RESULT 30
LOCUS BM454283 1035 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6402298 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5497394
5', mRNA sequence.
ACCESSION BM454283
VERSION BM454283.1 GI:18503323
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1035)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNP)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNP, at:
http://image.llnl.gov
Plate: LLAM12127 row: h column: 03
High quality sequence stop: 602.

```

FEATURES
  source
    1. 1035
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_image="5497394"
      /clone_lib="NIH_MGC_85"
      /tissue_type="lymphoma, cell line"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: Lymph: Vector: pCMV-SPORT6; Site: 1; Note:
      Average insert size 1.867 kb. Library enriched for
      full-length clones and constructed by Life Technologies.
      Note: this is a NIH_MGC Library."

BASE COUNT      284 a      226 c      263 g      261 t      1 others
ORIGIN

Query Match      2.8%; Score 21; DB 13; Length 1035;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 739 AAAATTTTATTTTATTTCT 759
|||||TTTTTTTTTTTTTTTTTTT
Db 981 AAAATTTTATTTTATTTCT 961

RESULT 31
BI070313      165 bp      mRNA      linear      EST 19-JUN-2001
LOCUS
DEFINITION    C017P98U Populus strain T89 leaves Populus tremula x Populus
               tremuloides cDNA, mRNA sequence.
ACCESSION     BI070313
VERSION       BI070313.1 GI:14490933
KEYWORDS
SOURCE
ORGANISM
  Populus tremula x Populus tremuloides.
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
  1 (bases 1 to 165)
  Jansson,S., Bhalero,R., Erlandsson,R., Bjorkbacka,H., Karlsson,J.,
  Steiky,F., Gustafsson,P. and Lundberg,J.
  Gene expression in Populus leaves
  Unpublished (2001)
  Contact: Erlandsson R
  Department of Biotechnology
  Royal Institute of Technology
  Teknikringen 30, Stockholm S-10044, Sweden
  Tel: 46 8 790 8287
  Fax: 46 8 245452
  Email: riker@biochem.kth.se.

FEATURES
  source
    1. 165
      /organism="Populus tremula x Populus tremuloides"
      /strain="T89"
      /db_xref="taxon:47664"
      /clone_lib="Populus strain T89 leaves"
      /tissue_type="leaf"

BASE COUNT      52 a      27 c      26 g      60 t
ORIGIN

Query Match      2.6%; Score 20; DB 13; Length 165;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 734 AAAATTTTATTTTATTTT 753
|||||TTTTTTTTTTTTTTTTTTT
Db 137 AAAATTTTATTTTATTTT 156

RESULT 32
AU074661      171 bp      mRNA      linear      EST 24-JUN-1999
LOCUS
DEFINITION    AU074661 Dictyostelium discoideum ss (H.Urushihara) Dictyostelium

```

```

ACCESSION      AU074661
VERSION        AU074661.1 GI:5181082
KEYWORDS
SOURCE
ORGANISM
  Dictyostelium discoideum.
  Dictyostelium discoideum
  Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
  1 (bases 1 to 171)
  Urushihara,H
  Developmental cDNA in Dictyostelium discoideum (1999)
  Unpublished (1999)
  Contact: Hideko Urushihara
  Institute of Biological Sciences
  University of Tsukuba
  1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
  Tel: 81-298-53-4664
  Fax: 81-298-53-6614
  Email: hideko@biol.tsukuba.ac.jp
  PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES
  source
    1. 171
      /organism="Dictyostelium discoideum"
      /strain="AX4"
      /db_xref="taxon:44689"
      /clone="SSL474"
      /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
      /dev_stage="slug"

BASE COUNT      69 a      7 c      1 g      94 t
ORIGIN

Query Match      2.6%; Score 20; DB 9; Length 171;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 738 TAAATTTTATTTTATTTT 757
|||||TTTTTTTTTTTTTTTTTTT
Db 49 TAAATTTTATTTTATTTT 68

RESULT 33
BI125558/C      222 bp      mRNA      linear      EST 31-DEC-2001
LOCUS
DEFINITION    I062P92P Populus leaf cDNA library Populus tremula x Populus
               tremuloides cDNA, mRNA sequence.
ACCESSION     BI125558
VERSION       BI125558.1 GI:18009529
KEYWORDS
SOURCE
ORGANISM
  Populus tremula x Populus tremuloides.
  Populus tremula x Populus tremuloides
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
  1 (bases 1 to 222)
  Hertzberg,M., Aspeborg,H., Erlandsson,R., Bjorkbacka,H., Hillonen
  ,T., Karlsson,J., Teeri,T., Gustafsson,P., Bhalero,R., Jansson,S.,
  Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and
  Lundberg,J.
  Gene expression in Populus
  Unpublished (2001)
  Contact: Erlandsson R
  Department of Biotechnology
  Royal Institute of Technology
  Teknikringen 30, Stockholm S-10044, Sweden
  Tel: 46 8 790 8287
  Fax: 46 8 245452
  Email: riker@biochem.kth.se.

FEATURES
  source
    1. 222
      /organism="Populus tremula x Populus tremuloides"
      /db_xref="taxon:47664"
      /clone_lib="Populus leaf cDNA library"
      /note="Organ: leaf"

BASE COUNT      138 a      18 c      48 t

```

ORIGIN

Query Match 2.6%; Score 20; DB 13; Length 222;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 738 TAAATTTTATTTATTT 757
|||||
Db 190 TAAATTTTATTTATTT 171

RESULT 34
LOCUS AU071341 240 bp mRNA linear EST 23-JUN-1999
DEFINITION AU071341 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSB257, mRNA sequence.
ACCESSION AU071341
VERSION AU071341.1 GI:5161530
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 240)
AUTHORS Urushihara,H.
TITLE Developmental cDNA in Dictyostelium discoideum (1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.
Location/Qualifiers
1..240
/organism="Dictyostelium discoideum"
/strain="Ax4"
/db_xref="taxon:44689"
/clone_lib="SSB257"
/clone="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 109 a 18 c 8 g 101 t 4 others

Query Match 2.6%; Score 20; DB 9; Length 240;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 738 TAAATTTTATTTATTT 757
|||||
Db 175 TAAATTTTATTTATTT 194

RESULT 35
LOCUS BM535454 248 bp mRNA linear EST 20-FEB-2002
DEFINITION EST588476 tomato breaker fruit Lycopersicon esculentum cDNA clone
cLE66023 5' end, mRNA sequence.
ACCESSION BM535454
VERSION BM535454.1 GI:18614121
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 248)
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
,J., Bougri,O., Kirkness,E., Uterback,T., Van Aken,S., Renning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)

JOURNAL Unpublished (2002)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
Location/Qualifiers
1..248
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLE66023"
/clone_lib="tomato breaker fruit"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="vector: pBluescriptSMCudapt; Site.1: EcoRI;
site.2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT 81 a 50 c 44 g 73 t

Query Match 2.6%; Score 20; DB 13; Length 248;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 ATTGAACAATCGAACAAT 179
|||||
Db 151 ATTGAACAATCGAACAAT 170

RESULT 36
LOCUS AV185077 300 bp mRNA linear EST 21-JUL-1999
DEFINITION AV185077 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk673c6 3', mRNA sequence.
ACCESSION AV185077
VERSION AV185077.1 GI:5564978
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 300)
Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A., Sano
,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,T. and
Nomoto,H.
Expressed genes in C.elegans
Unpublished (1999)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk673c6"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"

JOURNAL Unpublished (1999)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk673c6"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"


```

BASE COUNT      113 a      32 c      57 g      98 t
ORIGIN
Query Match      2.6%; Score 20; DB 9; Length 300;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 734 AAAATAAAATTTTATTTT 753
|||||
Db 29 AAAATAAAATTTTATTTT 10

RESULT 37
C35776 300 bp mRNA linear EST 18-OCT-1999
C35776/c
LOCUS
DEFINITION C35776 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
ACCESSION C35776
VERSION C35776.1 GI:2371917
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Pelodierinae; Caenorhabditis.
AUTHORS Kohara,Y., Motobashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111 Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
source
1.300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk448e12"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"

BASE COUNT 108 a 32 c 50 g 100 t 10 others
ORIGIN
Query Match 2.6%; Score 20; DB 14; Length 300;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 734 AAAATAAAATTTTATTTT 753
|||||
Db 29 AAAATAAAATTTTATTTT 10

RESULT 38
A2772185 324 bp DNA linear GSS 16-FEB-2001
A2772185/c
LOCUS
DEFINITION IM0574E24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0574E24 R, DNA sequence.
ACCESSION A2772185
VERSION A2772185.1 GI:12895229
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,

```

```

TITLE
JOURNAL
COMMENT
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0574 row: E column: 24
Seq primer: CACACAGCAACACACTATGACC
Class: plasmid ends
High quality sequence stop: 324.
FEATURES
location/Qualifiers
1..324
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0574E24"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g114732114[gb|AF129072.1]), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 114 a 60 c 43 g 107 t
ORIGIN
Query Match 2.6%; Score 20; DB 17; Length 324;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 737 ATAAATTTTATTTATTT 756
|||||
Db 134 ATAAATTTTATTTATTT 115

RESULT 39
A1356993 344 bp mRNA linear EST 15-FEB-1999
A1356993/c
LOCUS
DEFINITION qx15b09.x1 NCI-CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2001401 3'
similar to contains Alu repetitive element; mRNA sequence.
ACCESSION A1356993
VERSION A1356993.1 GI:4108614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Life Technologies catalog #: 11547-015
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1125 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 300.

FEATURES
 source
 1..344
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2001401"
 /clone_lib="NCI-CGAP_Lym12"
 /tissue_type="lymphoma, follicular mixed small and large
 cell"
 /lab_host="DH10B"
 /note="organ: lymph node; Vector: PCMV-SpORT6; Site_1:
 Salt; Site_2: NotI; Cloned unidirectionally. Primer:
 Oligo dt. Average insert size 1.25 kb. Life Technologies
 catalog #: 11547-015"
 68 a 81 c 84 g 111 t

BASE COUNT
ORIGIN

Query Match 2.6%; Score 20; DB 9; Length 344;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 738 TAAATTTTATTTTATTT 757
 ||||||||||||||||||
 Db 18 TAAATTTTATTTTATTT 37

RESULT 40 344 bp mRNA linear EST 30-NOV-2000
LOCUS BB579841 RIKEN full-length enriched, adult male adrenal gland Mus
DEFINITION BB579841 musculus cDNA 7330402A21 5', mRNA sequence.
ACCESSION BB579841
VERSION BB579841.1 GI:11475788
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 344)
 Alizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
 Hanagaki, T., Hayatsu, S., Hirooka, T., Hirose, T., Hodojima, Y.,
 Imotani, K., Ishii, Y., Itoh, M., Izawa, H., Kawai, J., Kojima, Y., Kono,
 H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K.,
 Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
 Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A.,
 Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka,
 T., Toya, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
 Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Alizawa, K. et al. 2000)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resesc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermostabilization and thermoactivation of thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kusunagi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.etc.riken.go.jp>) for
 further details.

FEATURES
 source
 1..344
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="7330402A21"
 /clone_lib="RIKEN full-length enriched, adult male adrenal
 gland"
 /sex="male"
 /tissue_type="adrenal gland"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Site_1: Salt; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5',
 GAGAGAGAGAGATCCCAAGACCTCTTTTCTTTTCTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 185.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5', GAGAGAGAGATCTCGAGTTATTAATTAATCCCCCCCCCC
 3']. cDNA was cloned into the XhoI and BamHI sites.
 Vector: a modified pBluescript KS(+) after bulk excision
 from Lambda FLC 1. Cloning sites, 5' end: Salt; 3' end:
 BamHI"

BASE COUNT 102 a 58 c 90 g 94 t

ORIGIN

Query Match 2.6%; Score 20; DB 10; Length 344;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 674 AAAAATTACCAATAGAC 693
 ||||||||||||||||||
 Db 277 AAAAATTACCAATAGAC 258

RESULT 41 391 bp mRNA linear EST 24-JUN-2002
LOCUS BO596136/6 B0596136
DEFINITION pfESTob29g01.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
 falciparum cDNA 5', mRNA sequence.
ACCESSION BO596136
VERSION BO596136.1 GI:21542862
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Hemosporida; Plasmodium.
 1 (bases 1 to 391)
 Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,
 Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theisling, B.,
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,
 Tsagarelashvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,
 Maguire, L., Ritchey, J., Wadkins, J., Kennedy, S., Levinso, D.,
 Waterston, R., Wilson, R. and Sibley, D.

TITLE Washu Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
Washu Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -400p from Gibco.
Location/Qualifiers

FEATURES

1..391

/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone.lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) - lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the poly(A)-tract
mRNA isolation system (Promega, WI) using streptavidin
magnisphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 Zapit vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the Exassist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

BASE COUNT 135 a 22 c 40 g 194 t
ORIGIN

Query Match 2.6%; Score 20; DB 14; Length 391;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 AATGAATTAATTTTATT 750

DB 304 AATGAATTAATTTTATT 285

RESULT 42

AZ823056/c

LOCUS AZ823056 403 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0096620R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0096620 R, DNA sequence.

ACCESSION AZ823056

VERSION AZ823056.1 GI:12992964

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 403)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0096 row: A column: 20
Seq primer: CACACAGAAACGCTATCACC
Class: plasmid ends
High quality sequence stop: 403.

FEATURES

source

1..403

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone.lib="UUGC2M0096620"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (G114732114191A1F19072.1), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapored mouse DNA was annealed to
adapored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 117 a 74 c 92 g 120 t
ORIGIN

Query Match 2.6%; Score 20; DB 17; Length 403;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 738 TAAATTTTATTTTATT 757

DB 214 TAAATTTTATTTTATT 195

RESULT 43

AU037304

LOCUS AU037304 409 bp mRNA linear EST 29-MAR-1999
DEFINITION AU037304 Dictyostelium discoideum SS (H.Urushi-hara) Dictyostelium
discoideum cDNA clone SSB739, mRNA sequence.

ACCESSION AU037304

VERSION AU037304.1 GI:3984057

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum

REFERENCE Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
1 (bases 1 to 409)

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,U., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.

TITLE The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development

JOURNAL DNA Res. 5 (6), 335-340 (1998)

COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tenoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664

Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
Location/Qualifiers
1. 409

FEATURES

/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone.lib="SLH156"

/dev_stage="slug"

BASE COUNT 152 a 42 c 12 g 200 t 3 others

ORIGIN

Query Match 2.6%; Score 20; DB 9; Length 409;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 738 TAAATTTTATTTTATTT 757
|||||

Db 79 TAAATTTTATTTTATTT 98
|||||

RESULT 44
A0039211/c 418 bp mRNA linear EST 29-MAR-1999

LOCUS A0039211 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium

DEFINITION discoidium cDNA clone SLH156, mRNA sequence.

ACCESSION A0039211

VERSION A0039211.1 GI:4008192

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Eukaryota; Eukaryota; Dictyostelidia; Dictyostelium.

REFERENCE 1 (bases 1 to 418)

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,

Yoshino,R., Mitra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,

Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.

The Dictyostelium development cDNA project: generation and

analysis of expressed sequence tags from the first-finger stage of

development

JOURNAL DNA Res. 5 (6), 335-340 (1998)

MEDLINE 99156227

COMMENT Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan

Tel: 81-298-53-4664

Fax: 81-298-53-6614

Email: hideko@biol.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.

BASE COUNT 202 a 90 c 33 g 93 t

ORIGIN

Query Match 2.6%; Score 20; DB 9; Length 418;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 738 TAAATTTTATTTTATTT 757
|||||

Db 404 TAAATTTTATTTTATTT 385
|||||

RESULT 45

B0576836/c

LOCUS B0576836 427 bp mRNA linear EST 19-JUN-2002

DEFINITION PFESTroab1004.y1 Plasmodium falciparum 307 asexual cDNA Plasmodium

ACCESSION B0576836

VERSION B0576836.1 GI:21480153

KEYWORDS EST.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 427)

AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,

Maria,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,

Bowers,T., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Konko,I.,

Tsagarisshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,

Maguire,L., Richey,J., Madkins,J., Kennedy,S., Levinso,D.,

Waterston,R., Wilson,R. and Sibley,D.

Washu Plasmodium EST Project

Unpublished (2001)

CONTACT: L. David Sibley

Washu Plasmodium EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University

Seq primer: -400p from Gibco.

FEATURES

source Location/Qualifiers

1. 427

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/clone.lib="Plasmodium falciparum 307 asexual cDNA"

/lab_host="DH10B (Genehelp, Invitrogen, Inc.)"

/note="Vector: pBluescript SK plus, Site1, EcoRI; Site2:

XhoI; Library was constructed by Debopam Chakrabarti.

Total RNA samples were isolated from mixed stage

saponin(0.1%) lysed P. falciparum 307 infected

erythrocytes by the acidic guanidium-phenol chloroform

method. The poly A+ RNA was isolated by the polyAT-Tract

mRNA isolation system (Promega, WI) using streptavidin

Magnosphere particles. Directional cDNA libraries were

constructed by oligo d(T) priming of poly(A)+ RNA (5mg)

into EcoRI and XhoI sites of 1 ZapII vector using the Zap

cDNA synthesis kit (Stratagene, CA). The average size of the

cDNA inserts in the library was between 1.0 and 1.5kb.

Clones were mass excised using the Exasist helper phage

(Stratagene), the phagemids were precipitated with PEG

8000 and extracted with phenol/chloroform. Phagemid DNA

was electroporated into DH10B cells."

BASE COUNT 165 a 27 c 49 g 186 t

ORIGIN

Query Match 2.6%; Score 20; DB 14; Length 427;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 AATAAATAAATTTTATTT 750
|||||

Db 414 AATAAATAAATTTTATTT 395
|||||

RESULT 46

B1713142/c 434 bp mRNA linear EST 11-MAR-2002

LOCUS B1713142

DEFINITION 1602604.y1 HR85 islet Homo sapiens cDNA clone IMAGE:5085895 5'

similar to SW:ALU8_HUMAN P39195 ALU SUBFAMILY SX SEQUENCE

CONTAMINATION WARNING ENTRY. [1] ;, mRNA sequence.

ACCESSION B1713142.1 GI:15688837

VERSION B1713142.1

KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 434)
 Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Hillier,L., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hamer,L., Marra,M., Page,D., Wylie,T., Martin,J., Blistein,A., Schmitt,A., Theising,B., Rittler,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarelisvill,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Other ESTs: ie02604.xl
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 424.
FEATURES
source
 Location/Qualifiers
 1..434
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5085895"
 /clone_lib="HR85 islet"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site:1; NotI; Site:2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@ingate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."
BASE COUNT
 90 a 109 c 86 g 148 t 1 others
ORIGIN
 Query Match 2.6%; Score 20; DB 13; Length 434;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 731 AATAAATTAATAATTTTATT 750
 ||||||||||||||||||
 Db 76 AATAAATTAATAATTTTATT 57
RESULT 47
 A0445188 451 bp DNA linear GSS 09-JAN-2001
LOCUS A0445188
DEFINITION GSSTC01459 Trypanosoma cruzi random genomic library Trypanosoma
ACCESSION A0445188
VERSION A0445188.3 GI:10131957
KEYWORDS GSS.
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi
 Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
REFERENCE
 1 (bases 1 to 451)
 Aguerro,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.
AUTHORS A random sequencing approach for the analysis of the trypanosoma
TITLE cruzi genome: general structure, large gene and repetitive DNA

JOURNAL families, and gene discovery
MEDLINE Genome Res. 10 (12), 1996-2005 (2000)
 20568489
COMMENT On Sep 14, 2000 this sequence version replaced gi:9371977.
 Contact: Sanchez D.O.
 Instituto de Investigaciones Biologicas (Univ. Nac. de Gral San Martin)
 Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24
 CP(1650) San Martin, Prov. de BS AS, Argentina
 Tel: 54-11-4580-7255
 Fax: 54-11-4752-9639
 Email: dsanchez@ib.unsam.edu.ar
 Sequences were basecalled with phred and vector was masked with crossmatch (see http://genome.washington.edu). Sequences were then trimmed from both ends to remove low quality bases and masked vector.
 Seq primer: 77
 Class: shotgun.
FEATURES
source
 Location/Qualifiers
 1..451
 /organism="Trypanosoma cruzi"
 /strain="Cl-Brener"
 /db_xref="taxon:5693"
 /clone="G1J19"
 /clone_lib="Trypanosoma cruzi random genomic library"
 /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was randomly sheared using a nebulizer and the 1 to 2 Kb range was gel purified and cloned into the dephosphorylated HincII site of the vector"
BASE COUNT
 172 a 145 c 54 g 80 t
ORIGIN
 Query Match 2.6%; Score 20; DB 17; Length 451;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 738 TAAATTTTATTTATTTATT 757
 ||||||||||||||||||
 Db 366 TAAATTTTATTTATTTATT 347
RESULT 48
 BH126751 458 bp DNA linear GSS 19-JUL-2001
LOCUS BH126751
DEFINITION BARC-Sat1585 Size-selected soybean genomic glycine max genomic, DNA
ACCESSION sequence.
 BH126751
VERSION BH126751.1 GI:14970254
KEYWORDS GSS.
SOURCE soybean.
ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 458)
 Cregan,P.B., Jarvik,T., Bush,A.L., Shoemaker,R.C., Lark,K.G.,
REFERENCE Kahler,A.L., Kaya,N., Vantol,T.T., Lohnes,D.G., Chung,J. and
AUTHORS Specht,J.E.
TITLE An integrated Genetic Linkage Map of the soybean genome
JOURNAL Crop Sci. 39, 1464-1490 (1999)
COMMENT Contact: Cregan PB
 Soybean Genomics and Improvement Lab
 USDA-ARS
 BARC-West, Bldg. 006, Beltsville, MD 20705, USA
 Tel: 301-504-5070
 Fax: 301-504-5728
 Email: cregan@ba.ars.usda.gov
 Single pass sequence. See the SoyBase home page
 (http://soybase.agron.iastate.edu) for PCR primer sequences and
 amplification conditions.
 Class: SSR-containing genome clone.

```

FEATURES
  source      Location/Qualifiers
              1..458
              /organism="Glycine max"
              /db_xref="taxon:3847"
              /clone.lib="Size-selected soybean genomic"
              /note="Simple sequence repeat containing clones from
              genomic DNA of the cultivar 'Williams' as described by
              Cregan, P.B., Bhagwat, A.A., Akaya, M.S. and Rongwen, J.
              (1994). Methods Cell. Mol. Biol. 5:49-61"

BASE COUNT      209 a      84 c      55 g      110 t
ORIGIN

Query Match      2.6%; Score 20; DB 17; Length 458;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      738 TAAATTTTATTTTATTT 757
        |||||||
Db      402 TAAATTTTATTTTATTT 383

RESULT 49
LOCUS      BG724967      459 bp      mRNA      linear      EST 08-MAY-2001
DEFINITION Etested3n04.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION  BG724967
VERSION     BG724967.1 GI:14008363
KEYWORDS   EST.
SOURCE      Eimeria tenella.
            Eimeria tenella.
ORGANISM    Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
            1 (bases 1 to 459)
REFERENCE   1 (bases 1 to 459)
AUTHORS    Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theisling,B., Allen
            M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
            WashU-Merck Eimeria tenella project
            Unpublished (1999)
TITLE       Contact: David Sibley, Ph.D.
JOURNAL     WashU-Merck Eimeria tenella project
COMMENT     Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@wustl.edu
            Contact David Sibley (toxoeast@orcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 428.

Location/Qualifiers
1..459
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone.lib="Eimeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of E. tenella grown in chickens. cDNA was
synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,
EcoRI adapters were ligated to the cDNA and products were
size-selected on sephacryl S500. The cDNA were ligated to
EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
converted to phagemids by mass excision using Exassist
helper phage and E.coli SOLR cell (Stratagene). Clones
were selected by negative hybridization against a pool of
overrepresented ESTs (N=10, from 1682 previous reads).
Insert sizes range from 1.2-2.9kb. The library may contain

```

```

BASE COUNT      164 a      41 c      65 g      188 t      1 others
ORIGIN

Query Match      2.6%; Score 20; DB 12; Length 459;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      734 AAATATAATTTTATTTT 753
        |||||||
Db      389 AAATATAATTTTATTTT 408

RESULT 50
LOCUS      BI439176      475 bp      mRNA      linear      EST 21-AUG-2001
DEFINITION ICG7d12.y3 HR85 islet Homo sapiens cDNA 5', mRNA sequence.
ACCESSION  BI439176
VERSION     BI439176.1 GI:15263866
KEYWORDS   EST.
SOURCE      human.
            human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 475)
REFERENCE   1 (bases 1 to 475)
AUTHORS    Melton,D., Brown,J., Kenty,G., Permut,A., Lee,C., Kaestner,K.,
            Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Rheising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            M., Gibbons,M., McCann,R., Cole,R., Tsagarisvill,R., Williams,T.,
            Jackson,Y. and Bowers,Y.
            Endocrine Pancreas Consortium
            Unpublished (2000)
TITLE       Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
JOURNAL     Endocrine Pancreas Consortium
COMMENT     Harvard University, Howard Hughes Medical Institute
            Department of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@hlp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center for information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            High quality sequence stop: 472.

Location/Qualifiers
1..475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="HR85 islet"
/tissue.type="Purified pancreatic islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site.1:
NotI; Site.2: XhoI; cDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size -1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permut Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."

BASE COUNT      100 a      116 c      106 g      152 t      1 others
ORIGIN

Query Match      2.6%; Score 20; DB 13; Length 475;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      731 AATATAATAATTTTATTT 750
        |||||||
Db      64 AATATAATAATTTTATTT 45

```

Search completed: February 20, 2003, 23:13:49
Job time : 1502 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 19:57:59 ; Search time 2207 Seconds
(without alignments)
10034.996 Million cell updates/sec

Title: US-10-024-955-6

Perfect score: 761

Sequence: 1 GATCTATATCAATACAAAT.....ATTTTATTTTATTTCTCC 761

Scoring table: IDENTITY_NUC
Gapex 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: GenEmbl:*
2: gb_ba:*
3: gb_hcg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	761	100.0	761	3 S80655	S80655 allergen De
2	761	100.0	761	6 AX028702	AX028702 Sequence
3	518.6	68.1	812	3 DP037044	U37044 Dermatopha
4	517	67.9	812	6 AX098569	AX098569 Sequence
5	517	67.9	812	6 AX028697	AX028697 Sequence
6	49.6	6.5	167614	2 AC119641	AC119641 Rattus no
7	48.2	6.3	108873	2 AC095814	AC095814 Rattus no
8	47.6	6.2	115758	9 AC104634	AC104634 Homo sapi
9	47.4	6.2	147671	2 AC105625	AC105625 Rattus no
10	46.6	6.1	236120	14 AF063866	AF063866 Melanopl
11	46.2	6.1	7921	3 AF153362	AF153362 Dictyoste
12	46.2	6.1	130760	14 AF303045	AF303045 Helicover
13	46.2	6.0	23154	2 AC115602	AC115602 Dictyoste
14	46	6.0	1141	6 AX083744	AX083744 Sequence
15	45.8	6.0	236542	2 AC096320	AC096320 Rattus no
16	45.2	5.9	169163	2 AC115666	AC115666 Rattus no
17	45.2	5.9	175119	2 AC105323	AC105323 Mus muscu
18	44.6	5.8	153477	2 AC006278	AC006278 Plasmodiu
19	44.4	5.8	14235	14 H2067264	H2067264 Helicover
20	44.2	5.8	130869	14 AF334030	AF334030 Helicover
21	44.2	5.8	131403	14 AF271059	AF271059 Helicove
22	44.2	5.8	172853	9 AC084361	AC084361 Homo sapi
23	44.2	5.8	107739	2 AC116979	AC116979 Dictyoste
24	44	5.8	165337	2 AC096212	AC096212 Rattus no
25	44	5.8	170668	2 AC091787	AC091787 Orzya sat
26	44	5.8	8813	2 AC115578	AC115578 Dictyoste
27	43.8	5.8	166613	2 AC099368	AC099368 Rattus no
28	43.8	5.8	195824	9 AL162726	AL162726 Human DNA
29	43.8	5.8	227145	2 AC096351	AC096351 Rattus no
30	43.8	5.8	239888	2 AC093939	AC093939 Rattus no
31	43.8	5.7	123078	8 AF004902	AF004902 Lotus jap
32	43.4	5.7	149228	2 AC113774	AC113774 Rattus no
33	43.4	5.7	164399	3 PFMA13P6	PFMA13P6 Plasmodium
34	43.4	5.7	110000	2 PFMA14P1_1	PFMA14P1_1 Continuatio
35	43.2	5.7	128656	2 AC124960	AC124960 Medicago
36	43.2	5.7	703	3 PCLRAP	X63619 P. chabaudi
37	43	5.7	30288	14 VV018338	VV018338 Varicella vir
38	43	5.7	110384	9 AC000117	AC000117 Homo sapi
39	43	5.7	217034	10 AC012147	AC012147 Mus muscu
40	43	5.7	219294	2 AC015797	AC015797 Mus muscu
41	43	5.7	231739	9 AC011811	AC011811 Homo sapi
42	43	5.7	259940	2 AC115288	AC115288 Mus muscu
43	43	5.6	14261	2 AC116031	AC116031 Dictyoste
44	42.8	5.6	152082	2 AL807792	AL807792 Danio rer
45	42.8	5.6	152082	2	

ALIGNMENTS

RESULT 1
LOCUS: S80655 761 bp. mRNA linear INV 10-MAY-1996
DEFINITION: allergen Der f 7 [Dermatophagoides pteronyssinus-house dust mites,
mRNA, 761 nt].
ACCESSION: S80655
VERSION: S80655.1 GI:1311688
KEYWORDS: ACCESSION
SOURCE: S80655.1
ORGANISM: Dermatophagoides pteronyssinus.
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariiformes; Sarcophiliiformes; Astigmata; Analgoidea; Pyroglyphidae;
Dermatophagoides.
REFERENCE
1 (bases 1 to 761)
Shen,H.D., Chua,K.Y., Lin,W.L., Hsieh,K.H. and Thomas,W.R.

TITLE Molecular cloning and immunological characterization of the house dust mite allergen Der f 7
JOURNAL Clin. Exp. Allergy 25 (10), 1000-1006 (1995)
MEDLINE 96120794
PUBMED 8556554
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI g14942] from the original journal article.
This sequence comes from Fig. 1.
location/Qualifiers
1..761
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
1..761
/gene="allergen Der f 7"
43..684
/note="This sequence comes from Fig. 1"
/codon_start=1
/product="allergen Der f 7"
/protein_id="AAB35977.1"
/db_xref="GI:1311689"
/translation="MMKFLIIAAVAVFVNSADPIHYDKITEENKAIDAIAIIEOSE
TIDPMKVPDHADKFERHVCIVDPKGEIARNRLEAKLQMKRQGDANVGEISYRAH
LILGVHDDIVSMEDLAKLGDHPHTTHVLSIDIVVALSLSEISDEGNITMTSEVR
QFANVNVNHIGSLIDPIFGVLSDVLTAFODTVREKMTKVLAPARKELEKN"
BASE COUNT 245 a 134 c 137 g 245 t
ORIGIN
Query Match 100.0%; Score 761; DB 3; Length 761;
Best Local Similarity 100.0%; Pred. No. 5.8e-136;
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCTTATATCAATTAACATCAAAAAACATATCTTACAAAATGATGAATTTTGTG 60
DB 1 GATCTTATATCAATTAACATCAAAAAACATATCTTACAAAATGATGAATTTTGTG 60
QY 61 ATTGTCGCGTGCAATTTGTGCGCGTTTGGCGTATCAATCAATCATATGATAAATCACC 120
DB 61 ATTGTCGCGTGCAATTTGTGCGCGTTTGGCGTATCAATCAATCATATGATAAATCACC 120
QY 121 GAAGAATCAACAAGTATGATGATCCATGCTGCTCTATTGAACAATCCGAACAATA 180
DB 121 GAAGAATCAACAAGTATGATGATCCATGCTGCTCTATTGAACAATCCGAACAATA 180
QY 181 GATCCAAATGAAGTACATGATCCGATTAATTCGAACGTCTGTGATTTGGAT 240
DB 181 GATCCAAATGAAGTACATGATCCGATTAATTCGAACGTCTGTGATTTGGAT 240
QY 241 TTCAAGGTGAATTTAGCCATGCGAATTTGAGGCTCGAGGATTTGAACAATGAACGT 300
DB 241 TTCAAGGTGAATTTAGCCATGCGAATTTGAGGCTCGAGGATTTGAACAATGAACGT 300
QY 301 CAAGGTGATGCTAATGCAAGGTGAAGAGGTATTGTAAAGTCAATTTGTGATCGGT 360
DB 301 CAAGGTGATGCTAATGCAAGGTGAAGAGGTATTGTAAAGTCAATTTGTGATCGGT 360
QY 361 GTTCACGATATATCGCTCGATGAGATATGATTTAGCATACAAATGGGATCTTCAT 420
DB 361 GTTCACGATATATCGCTCGATGAGATATGATTTAGCATACAAATGGGATCTTCAT 420
QY 421 CCAACCACTCATGTCATTTGGATATTCAAGATTTTGTGTGCGCTTGCGCTTGAATTT 480
DB 421 CCAACCACTCATGTCATTTGGATATTCAAGATTTTGTGTGCGCTTGCGCTTGAATTT 480
QY 481 TTCGATGAAGGTAAATGAATGACATCTTTTGAATAGACAAATTCGCTAAAGTTGTC 540
DB 481 TTCGATGAAGGTAAATGAATGACATCTTTTGAATAGACAAATTCGCTAAAGTTGTC 540
QY 541 AACCATATTGGTGTCTTTCAATCTTGATCAATTTTGGCGTTTATCTGATGATTTG 600
DB 541 AACCATATTGGTGTCTTTCAATCTTGATCAATTTTGGCGTTTATCTGATGATTTG 600
QY 601 ACCGCTATTTTCCAAAGACCGTACGTAAGAAATGACCAAAATGATTGGACACGATTT 660

|||||
DB 601 ACCGCTATTTTCCAAAGACCGTACGTAAGAAATGACCAAAATGATTGGACACGATTT 660
QY 661 AAACGTGAATTTGAAAAAATTTAACCAATGACATCATTTTCCAACTGACAAATCTCA 720
DB 661 AAACGTGAATTTGAAAAAATTTAACCAATGACATCATTTTCCAACTGACAAATCTCA 720
QY 721 TTTCACGTACAAATTAATAATTTTATTTATTTATTTCTCC 761
DB 721 TTTCACGTACAAATTAATAATTTTATTTATTTATTTCTCC 761
RESULT 2
AX028702 761 bp DNA linear PAT 16-SEP-2000
LOCUS AX028702
DEFINITION Sequence 6 from Patent EP1018550.
ACCESSION AX028702
VERSION AX028702.1 GI:10189815
KEYWORDS European house dust mite.
SOURCE Dermatophagoides pteronyssinus
ORGANISM Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Acariformes; Sarcoptiformes; Astigmata; Anagloidea; Pyroglyphidae; Dermatophagoides.
REFERENCE 1 (bases 1 to 761)
AUTHORS Thomas W.R. and Chua K.Y.
TITLES Allergenic protein and peptides from house dust mite and uses thereof
JOURNAL Patent: EP 1018550-A 6 12-JUL-2000;
INST CHILD HEALTH RESEARCH (AU)
FEATURES
source
1..761
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
43..684
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09235.1"
/db_xref="GI:10189816"
CDS
/translation="MMKFLIIAAVAVFVNSADPIHYDKITEENKAIDAIAIIEOSE
TIDPMKVPDHADKFERHVCIVDPKGEIARNRLEAKLQMKRQGDANVGEISYRAH
LILGVHDDIVSMEDLAKLGDHPHTTHVLSIDIVVALSLSEISDEGNITMTSEVR
QFANVNVNHIGSLIDPIFGVLSDVLTAFODTVREKMTKVLAPARKELEKN"
BASE COUNT 245 a 134 c 137 g 245 t
ORIGIN
Query Match 100.0%; Score 761; DB 6; Length 761;
Best Local Similarity 100.0%; Pred. No. 5.8e-136;
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCTTATATCAATTAACATCAAAAAACATATCTTACAAAATGATGAATTTTGTG 60
DB 1 GATCTTATATCAATTAACATCAAAAAACATATCTTACAAAATGATGAATTTTGTG 60
QY 61 ATTGTCGCGTGCAATTTGTGCGCGTTTGGCGTATCAATCAATCATATGATAAATCACC 120
DB 61 ATTGTCGCGTGCAATTTGTGCGCGTTTGGCGTATCAATCAATCATATGATAAATCACC 120
QY 121 GAAGAATCAACAAGTATGATGATCCATGCTGCTCTATTGAACAATCCGAACAATA 180
DB 121 GAAGAATCAACAAGTATGATGATCCATGCTGCTCTATTGAACAATCCGAACAATA 180
QY 181 GATCCAAATGAAGTACATGATCCGATTAATTCGAACGTCTGTGATTTGGAT 240
DB 181 GATCCAAATGAAGTACATGATCCGATTAATTCGAACGTCTGTGATTTGGAT 240
QY 241 TTCAAGGTGAATTTAGCCATGCGAATTTGAGGCTCGAGGATTTGAACAATGAACGT 300
DB 241 TTCAAGGTGAATTTAGCCATGCGAATTTGAGGCTCGAGGATTTGAACAATGAACGT 300
QY 301 CAAGGTGATGCTAATGCAAGGTGAAGAGGTATTGTAAAGTCAATTTGTGATCGGT 360
DB 301 CAAGGTGATGCTAATGCAAGGTGAAGAGGTATTGTAAAGTCAATTTGTGATCGGT 360

FEATURES	source
CD5	
LOCUS	DPU37044
DEFINITION	812 bp mRNA linear INV 31-OCT-1995
ACCESSION	Dermatophagoides pteronyssinus Der p 7 allergen polypeptide mRNA,
VERSION	complete cds.
KEYWORDS	U37044
SOURCE	U37044.1 GI:1045601
ORGANISM	Dermatophagoides pteronyssinus.
REFERENCE	Dermatophagoides pteronyssinus
AUTHORS	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
TITLE	Acariiformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;
	Dermatophagoides.
	1 (bases 1 to 812)
	Shen, H.D., Chua, K.Y., Lin, K.L., Hsieh, K.H. and Thomas, W.R.
	Molecular cloning of a house dust mite allergen with common
	antibody binding specificities with multiple components in mite
	extracts
JOURNAL	Clin. Exp. Allergy 23 (11), 934-940 (1993)
MEDLINE	20239583
PUBMED	10779281
REFERENCE	2 (bases 1 to 812)
TITLE	Thomas, W.R.
JOURNAL	Direct Submission
	Submitted (26-SEP-1995) Wayne R. Thomas, Molecular Biology
	Division, Institute for Child, Health Research, Princess Margaret
	Hospital, Roberts Road, Subiaco, Western Australia, 6008, Australia
	location/Qualifiers
	1. 812
	/organism="Dermatophagoides pteronyssinus"
	/db_xref="taxon:6956"
	/clone="HD6"
	68..715
	/codon_start=1
	/product="Der p 7 allergen polypeptide"
	/protein_id="AA80264.1"
	/db_xref="GI:1045602"
	/translation="MKRLILAAAFVAVSADPIHYHKTPEIRKAVDEAVAIETKSE
	TTPDQKVDHSDKREHRIYGIIDLGEMINIOYRGKOKMRGVDAVNSDDGVKRAH
	LIVCGHDVVSMEDVLA YKIGDLHPNHVSIIDDPVVELSTLSEYSGNMRTSFEYR
	QFANVNVHIGSLIDPIPAVLSVYLAIRPDYRAEKTVLAFAFKELERNQ"
	68..118
	/note="putative leader peptide"

Query Match	68.1%	Score 518.6	DB 3	Length 812
Best Local Similarity	81.4%	Pred. No. 1.5e-89		
Matches 613; Conservative	0	Mismatches 139	Indels 1	Gaps 1

OY	3	TCCTTATTCATTAACAAATCCAAAAAAACAATFATCTTACAAATGATGAATTTTGGTAT	62
Db	28	TTTTTCATATCGTAAAAATCCAAATTCACITTTTTTACCAAAATGATGAATTTATTTGAT	87
OY	63	TGCTGCGGTGGCATTTGTCGCCGTTTTCGGCTGATCCAAATTCATATGATTAATAATCACC	122
Db	88	TGCTGCGCGAGTTTGTGTTGGCGTTTGGCGTATCCAAATTCATATGATTAATAATCACC	147
OY	123	AGAAATCAACAAGCTATTTGATGATGCCATTGCTGCTATTGACAACTCCGAACATAGA	182
Db	148	AGAAATTAACAAAGCTGTTGATGAAGCCGTCGTCGAATTAATAAATCCGAACATTCGA	207
OY	163	TCCCATGGAAGACCTGATATGCCGATTAATAATGCAAGCTATGTTGGTATTTGGATTT	242
Db	208	TCCAAATGAAGGATCCCGATCATCTGATTAATAATGCAAGACATATTTGGTATATCGATTT	267
OY	243	CAAAAGTAAATTTAGCCATGCGAAACATTTGAGGCTCGAGATTTGAACAAATGAACGTC	302
Db	268	AAAAGTGATTTAGACATATCGCAACATTCACAGTTGAGAGATTAACAAATGAACGTC	327
OY	303	AGGTGATCTTAATGTCAAAGGTGAAGGATATTTGTTAAAGCTCATTTTGTGATCGGT	362
Db	328	AGGTATGCTAATGTAAGAAAGTGAAGGATGTTGTGTCAAAGCTCATTTGTGTCGGGT	387
OY	363	TCACATATATCGCTCGATGAGATATGATTTAGCATACAAATTTGGGTGATCTTCATCC	422
Db	388	TCATATGATACGTTGTTCAATGGAATATGATTTAGCAATCAAAATTTGGGTGATCTTCATCC	447
OY	423	AACCACTGATGTCATTTGCGATATTCAGAAATTTTGTGTCCTGTGCTTGAATTTTC	482
Db	448	AAACGCTATGTCATTTGCGAATATTCAGAGATTTTGTGTCGAATATGCTCGGAAGTTAG	507
OY	483	TGATGAAGTAAACATACAAATGACATCTTTTGAAGTACGACAAATTCGCTAATGTTGCA	542
Db	508	CGAAGAGGATTAATGACATTTGACATCGTTCGAAGTACGTCAATTTGCCAATGTTGTC	567
OY	543	CCATATTTGGTGTCTTTCAATCTTGGATTCCAATTTTGGCTTTTATATGATGATTCAC	602
Db	568	TCATATTTGGTGTCTTTCAATTTTGGATTCCAATTTTGGCTTTTATATGATGATTTGAC	627
OY	603	CGCTATTTTCCAAAGCACCGTACGTAAGGAATGACCAAAATTTTGGGACACAGATTTAA	662
Db	628	CGCTATTTTCCAAAGTACCGTACGTAAGGAATGACCAAAATTTTGGGACACAGATTTAA	687
OY	663	ACGTGAATTTGAAAAAATTAACCAATAGACATCATTTTTCOAACGTACAACTCTATT	722
Db	688	AAAAAATTTGGAACGAA -CAACCAATAGACTTTACACACAAACATATACCTGTTATTTTA	746
OY	723	TCATCGACATTAATAATTTTATTTTTAT	755
Db	747	CACGTGATATCAAAATGAATAATAATTTTAT	779

RESULT 4			
AR098569			
LOCUS	812 bp	DNA	linear
DEFINITION	Sequence 1 from patent US 6077517.		PAT 14-FEB-2001
ACCESSION	AR098569		
VERSION	AR098569.1	GI:12808335	
KEYWORDS			
SOURCE	Unknown.		

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas W. Robert, and Chua, K.-Y.
TITLE House dust mite allergen, Der p VII, and uses thereof
JOURNAL Patent: US 6077517-A 1 20-JUN-2000;
FEATURES
source
1. 812
Location/Qualifiers
BASE COUNT 270 a 133 c 142 g 267 t
ORIGIN
Query Match 67.9%; Score 517; DB 6; Length 812;
Best Local Similarity 81.3%; Pred. No. 3e-89;
Matches 612; Conservative 0; Mismatches 140; Indels 1; Gaps 1;
QY 3 TCTTATATCAATTAACATCCAAAAAATATCTTACAAATGATGAATTTTGTGAT 62
DB 28 TTTTTCATATCGTAAATCCAAATTCACCTTTTACCAATGATGAATTAATTTATGAT 87
QY 63 TGGTCCGCGGCAATTTGCGCGCTTGGCTGATCCAAATTCAGATGATTAATTCACCGA 122
DB 88 TGTGCGCGGCGCTTTTGTCCGCTTGGCTGATCCAAATTCAGATGATTAATTCACCGA 147
QY 123 AGAATTCACAAAGCTATGATGATGCGCATTTGCTGATTTGAACCAATCCGAACATAGA 182
DB 148 AGAATTCACAAAGCTGATGATGACCGCTGCGCAATTCGAAATCCGAACATTCGA 207
QY 183 TCCAATGAAGTACGTCGATGATGCGCATTAATTCGAACGTCATGTTGTTGGATTT 242
DB 208 TCCAATGAAGTACCGCATGATTTCTGATTAATTCGAACGTCATGTTGTCATCGATTT 267
QY 243 CAAAGTGATTTAGCCATCGGAAACATTTGAGGCTCGAGATTGAACAAATGAAGCTCA 302
DB 268 AAAAGTCATTTAGACATGGGAAACATTTCAAGTTGCGAGATTAAACAAATGAAGCTGT 327
QY 303 AGGTGATGCTAATGCTCAAGGTGAGAGGATTTGTTAAAGCTCATTTGTTGATCGGTGT 362
DB 328 AGGTGATGCTAATGCTCAAGGTGAGAGGATTTGTTAAAGCTCATTTGTTGATCGGTGT 387
QY 363 TCAAGATGATGCTGCTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 422
DB 388 TCAATGATGACGTTGTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 447
QY 423 AACCATCATGCTATTTGCGATATTCAGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 482
DB 448 AAACACATCATGCTATTTGCGATATTCAGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 507
QY 483 TGATGAAGTAAACATTAACATGATCTTTTGAAGTACGACAAATTCGCTAATGTTGCA 542
DB 508 CGAAGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 567
QY 543 CCAATATGCTGCTTTCATATTCGATTCGAATTTTGGGGTATTATCTGATGATGATGATG 602
DB 568 TCAATATGCTGCTTTCATATTCGATTCGAATTTTGGGGTATTATCTGATGATGATGATG 627
QY 603 CGGTATTTTTCAGACACCGCTAGTGAAGAAATGACCAAGTATTTGACACAGCATTTAA 662
DB 628 CGGTATTTTTCAGACACCGCTAGTGAAGAAATGACCAAGTATTTGACACAGCATTTAA 687
QY 663 ACGTGAATGGAAGAAATTAACCAATGATGATCTTTTTCACCACTGATGATGATGATG 722
DB 688 AAAAGATTTGGAAGAAATTAACCAATGATGATCTTTTTCACCACTGATGATGATGATG 746
QY 723 TCACTGATCAATTAATTAATTTTATTTTAT 755
DB 747 CACTGATTAATCAATTAATTAATTTTAT 779

AX028697
VERSION AX028697.1 GI:10189810
KEYWORDS
SOURCE European house dust mite.
ORGANISM Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcoptiformes; Astigmata; Analoidea; Pyroglyphidae;
Dermatophagoides.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas W.R. and Chua, K.Y.
TITLE Allergenic protein and peptides from house dust mite and uses thereof
JOURNAL Patent: EP 1018550-A 1 12-JUL-2000;
INST CHILD HEALTH RESEARCH (AU)
FEATURES
source
1. 812
Location/Qualifiers
CDS
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
68..715
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09234.1"
/db_xref="GI:10189811"
/translation="MKLLILIAAFAVAVSADPIHYDKITERINKAVDVAIAIKSE
TPDPKVPDSDKFERHIGIDILKQDLNRNIOVGLQMKRVGDAVNSLEDGVKHA
LIVGVHDVVSMEYLAIVKLDIHNQTHIISIDIOFVELSLVESEGMLTSEVR
QFANVNHIGLSILDPFPAVLSVLTALFQTVRAEMKVLPAFKKELERNQ"
mat.peptide 119
BASE COUNT 270 a 133 c 142 g 267 t
ORIGIN
Query Match 67.9%; Score 517; DB 6; Length 812;
Best Local Similarity 81.3%; Pred. No. 3e-89;
Matches 612; Conservative 0; Mismatches 140; Indels 1; Gaps 1;
QY 3 TCTTATATCAATTAACATCCAAAAAATATCTTACAAATGATGAATTTTGTGAT 62
DB 28 TTTTTCATATCGTAAATCCAAATTCACCTTTTACCAATGATGAATTAATTTATGAT 87
QY 63 TGGTCCGCGGCAATTTGCGCGCTTGGCTGATCCAAATTCAGATGATTAATTCACCGA 122
DB 88 TGTGCGCGGCGCTTTTGTCCGCTTGGCTGATCCAAATTCAGATGATTAATTCACCGA 147
QY 123 AGAATTCACAAAGCTATGATGATGCGCATTTGCTGATTTGAACCAATCCGAACATAGA 182
DB 148 AGAATTCACAAAGCTGATGATGACCGCTGCGCAATTCGAAATCCGAACATTCGA 207
QY 183 TCCAATGAAGTACCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 242
DB 208 TCCAATGAAGTACCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 267
QY 243 CAAAGTGATTTAGCCATCGGAAACATTTGAGGCTCGAGATTGAACCAATGAAGCTCA 302
DB 268 AAAAGTCATTTAGACATGGGAAACATTCAGTTGCGAGATTAAACAAATGAAGCTGT 327
QY 303 AGGTGATGCTAATGCTCAAGGTGAGAGGATTTGTTAAAGCTCATTTGTTGATCGGTGT 362
DB 328 AGGTGATGCTAATGCTCAAGGTGAGAGGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 387
QY 363 TCAAGATGATGCTGCTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 422
DB 388 TCAATGATGACGTTGTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 447
QY 423 AACCATCATGCTATTTGCGATATTCAGATTTTGGTTGTTGTTGTTGTTGTTGTTGTTGTTG 482
DB 448 AAACACATCATGCTATTTGCGATATTCAGATTTTGGTTGTTGTTGTTGTTGTTGTTGTTGTTG 507
QY 483 TGATGAAGTAAACATTAACATGATCTTTTGAAGTACGACAAATTCGCTAATGTTGCA 542
DB 508 CGAAGAGTAAATTAACATTAACATGATCTTTTGAAGTACGACAAATTCGCTAATGTTGCA 567
QY 543 CCAATATGCTGCTTTCATATTCGATTCGAATTTTGGGGTATTATCTGATGATGATGATG 602

Db 568 TCATATGTCGTCTTTCATTTGGATTCATTTTCGTCGTCTTATCCGATGTTTTCAGC 627
Qy 603 CGCATTTTCCAGACACCGTACGTACGAAATGACCAAGATTTGGACCGACCATTTAA 662
Db 628 CGCATTTTCCAGATACCGTACGTACGAAATGACCAAGATTTGGACCGACCATTTAA 687
Qy 663 ACCTGAATTTGAAAAAATTAACCAATGACATCTATTTTCCACTGATACATCTCTATT 722
Db 688 AAAAGAAATTTGAGCAAAA-CAACCAATGACTTACACACACATACACTGTATTATTTTAA 746
Qy 723 TCACGTACAAATTAATTAATTTTATTTTAT 755
Db 747 CACTGGAATATCAATGAATGAATTTTAT 779

RESULT 6
AC119641/c 167614 bp DNA linear HTG 18-JUL-2002
LOCUS Rattus norvegicus clone CH230-428C17, *** SEQUENCING IN PROGRESS
DEFINITION *** 24 unordered pieces.
AC119641
AC119641.5 GI:21747111
VERSION HTG: HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 167614)
Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonni,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Butcher,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Dayla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denny,A.L., Ding,Y., Dinh,H.H.,
Douthaite,K.J., Draper,H., Dugan-Kocha,S., Durbin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsl,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheiswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savary,G.,
Schier,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Soderren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swatek,A., Tabak,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
TITLE Unpublished
JOURNAL 2 (bases 1 to 167614)
REFERENCE
AUTHORS Worley,K.C.

Direct Submission
Submitted (30-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 167614)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20467815.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: GVNN
Center clone name: CH230-428C17
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 151460 bases at least Q40
Consensus quality: 154539 bases at least Q30
Consensus quality: 156831 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1244: contig of 1244 bp in length
* 1245 1344: gap of unknown length
* 1345 2615: contig of 1271 bp in length
* 2616 6063: gap of unknown length
* 6064 6163: contig of 3348 bp in length
* 6164 8014: gap of unknown length
* 8015 8114: contig of 1851 bp in length
* 8115 9756: gap of unknown length
* 9759 9858: contig of 1644 bp in length
* 9859 11393: gap of unknown length
* 11394 14940: contig of 1535 bp in length
* 14941 15040: gap of unknown length
* 15041 18422: contig of 3447 bp in length
* 18423 18522: gap of unknown length
* 18523 21432: contig of 3382 bp in length
* 21433 25138: gap of unknown length
* 25139 25238: contig of 2810 bp in length
* 25239 31200: gap of unknown length
* 31201 37046: contig of 3706 bp in length
* 37047 37146: gap of unknown length
* 37147 42617: contig of 5471 bp in length
* 42618 49011: gap of unknown length
* 49012 49111: contig of 6294 bp in length
* 49112 54511: gap of unknown length
* 54512 54611: contig of 5400 bp in length
* 54612 62606: gap of unknown length
* 62607 62706: contig of 7995 bp in length
* 62707 71607: gap of unknown length
* 71608 71707: contig of 8901 bp in length
* 71708 79834: gap of unknown length
* 79835 79934: contig of 8127 bp in length
* 79935 gap of unknown length

```

* 79935 89709: contig of 9775 bp in length
* 89710 89809: gap of unknown length
* 89810 99550: contig of 9741 bp in length
* 99551 99650: gap of unknown length
* 99651 110776: contig of 11126 bp in length
* 110777 110876: gap of unknown length
* 110877 127046: contig of 16170 bp in length
* 127047 127146: gap of unknown length
* 127147 142799: contig of 15653 bp in length
* 142800 142899: gap of unknown length
* 142900 167614: contig of 24715 bp in length.
FEATURES
    source          Location/Qualifiers
1..167614          /organism="Rattus norvegicus"
                   /db_xref="taxon:10116"
                   /clone="CH230-428C17"
BASE COUNT      49648 a 32715 c 32706 g 48967 t 3578 others
ORIGIN
Query Match      6.5%; Score 49.6; DB 2; Length 167614;
Best Local Similarity 48.0%; Pred. No. 4.1;
Matches 142; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 304 GCGATGCTTAATGTCGAAGTGAAGGCGATGTTAAAGCTCATTTGTCATCGGTGT 363
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 13826 GTTGATGATGATGATGTTGTTGTTGATGTTGTTGTTGTTGTTGTTGTTGTT 13767
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 364 CACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 13766 GTTGCTGATATTTGTCATGTTGTTGTCATGTTGTCATGTTGTTGTTGTCATGTT 13707
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 424 ACCACATGTCATTTGCGATATTCAGATTTGTTGTCCTTGCCTTCGTAATTTCT 483
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 13706 GTTGTTGTTGTTGTTGTCGATGTTGTTGTCATGTTGTTGTTGTTGTCATGTTGTT 13647
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 484 GATGAAGGTACATTAACAATGACATCTTTTGAAGTACGACAAATTCGTAATGTTGTCAC 543
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 13646 GATGTTGTTGTTGATGTTGTTGTTGTTGATGTTGTTGTTGTTGTTGTTGTTGTT 13587
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 544 CATATTGTCGCTTCATCTTCGATCCAAATTTTGGCTTATCTGATGATTT 599
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 13586 GATGTTGTTGTTGTTGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 13531
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 7
AC111217 196650 bp DNA linear HTG 13-JUL-2002
LOCUS      Rattus norvegicus clone CH230-62D9, *** SEQUENCING IN PROGRESS ***,
DEFINITION 57 unordered pieces.
ACCESSION AC111217
VERSION    AC111217.2 GI:21735754
KEYWORDS  HTG: HTGS PHASE1.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
1 (bases 1 to 196650)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbarta,J., Benton,J., Blimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devalla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doubtwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabelli,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

```

```

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudh,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Kovach,J., Kover,C.,
Kratovic,U., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moran,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwuonu,G.,
Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,R., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshlari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 196650)
Worley K.C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 196650)
Worley K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18700980.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GI02
Center clone name: CH230-62D9
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 160704 bases at least Q40
Consensus quality: 165592 bases at least Q30
Consensus quality: 170308 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 1292 1291: contig of 1291 bp in length
* 1392 1391: gap of unknown length
* 1392 2876: contig of 1485 bp in length
* 2877 2976: gap of unknown length
* 2977 4164: contig of 1188 bp in length
* 4165 4264: gap of unknown length
* 4265 5344: contig of 1080 bp in length

```

* 5345 5444: gap of unknown length
* 5445 7029: contig of 1585 bp in length
* 7030 7129: gap of unknown length
* 7130 9421: contig of 2292 bp in length
* 9422 9521: gap of unknown length
* 9522 11094: contig of 1573 bp in length
* 11095 11194: gap of unknown length
* 11195 12281: contig of 1087 bp in length
* 12282 12381: gap of unknown length
* 12382 13992: contig of 1611 bp in length
* 13993 14092: gap of unknown length
* 14093 15607: contig of 1515 bp in length
* 15608 15707: gap of unknown length
* 15708 18296: contig of 2589 bp in length
* 18297 18396: gap of unknown length
* 18397 19808: contig of 1412 bp in length
* 19809 19908: gap of unknown length
* 19909 21168: contig of 1260 bp in length
* 21169 21268: gap of unknown length
* 21269 22646: contig of 1378 bp in length
* 22647 22746: gap of unknown length
* 22747 24805: contig of 2059 bp in length
* 24806 24905: gap of unknown length
* 24906 26845: contig of 1940 bp in length
* 26846 26945: gap of unknown length
* 26946 28488: contig of 1543 bp in length
* 28489 28588: gap of unknown length
* 28589 30084: contig of 1496 bp in length
* 30085 30184: gap of unknown length
* 30185 31721: contig of 1537 bp in length
* 31722 31821: gap of unknown length
* 31822 34036: contig of 2215 bp in length
* 34037 34136: gap of unknown length
* 34137 36808: contig of 2672 bp in length
* 36809 36908: gap of unknown length
* 36909 38581: contig of 1673 bp in length
* 38582 38681: gap of unknown length
* 38682 41373: contig of 2692 bp in length
* 41374 41473: gap of unknown length
* 41474 43808: contig of 2335 bp in length
* 43809 43908: gap of unknown length
* 43909 46254: contig of 2346 bp in length
* 46255 46354: gap of unknown length
* 46355 48995: contig of 2641 bp in length
* 48996 49095: gap of unknown length
* 49096 51734: contig of 2639 bp in length
* 51735 51834: gap of unknown length
* 51835 53932: contig of 2098 bp in length
* 53933 54032: gap of unknown length
* 54033 55674: contig of 1642 bp in length
* 55675 55774: gap of unknown length
* 55775 58807: contig of 3033 bp in length
* 58808 58907: gap of unknown length
* 58908 61237: contig of 2330 bp in length
* 61238 61337: gap of unknown length
* 61338 64863: contig of 3526 bp in length
* 64864 64963: gap of unknown length
* 64964 68383: contig of 3420 bp in length
* 68384 68483: gap of unknown length
* 68484 72066: contig of 3583 bp in length
* 72067 72166: gap of unknown length
* 72167 75513: contig of 3347 bp in length
* 75514 75613: gap of unknown length
* 75614 78382: contig of 2669 bp in length
* 78383 78382: gap of unknown length
* 78383 81978: contig of 3596 bp in length
* 81979 82078: gap of unknown length
* 82079 84178: contig of 2100 bp in length
* 84179 84278: gap of unknown length
* 84279 88697: contig of 4419 bp in length
* 88698 88797: gap of unknown length
* 88798 92474: contig of 3677 bp in length
* 92475 92574: gap of unknown length

* 92575 97329: contig of 4755 bp in length
* 97330 97429: gap of unknown length
* 97430 101714: contig of 4285 bp in length
* 101715 101814: gap of unknown length
* 101815 106144: contig of 4330 bp in length
* 106145 106244: gap of unknown length
* 106245 112003: contig of 5759 bp in length
* 112004 112103: gap of unknown length
* 112104 116236: contig of 4133 bp in length
* 116237 116336: gap of unknown length
* 116337 119380: contig of 3044 bp in length
* 119381 119480: gap of unknown length
* 119481 123749: contig of 4269 bp in length
* 123750 123849: gap of unknown length
* 123850 129497: contig of 5648 bp in length
* 129498 129597: gap of unknown length
* 129598 135379: contig of 5782 bp in length
* 135380 135479: gap of unknown length
* 135480 141999: contig of 6520 bp in length
* 142000 142099: gap of unknown length
* 142100 146302: contig of 4203 bp in length
* 146303 146402: gap of unknown length
* 146403 152138: contig of 5736 bp in length
* 152139 152238: gap of unknown length
* 152239 160905: contig of 8667 bp in length
* 160906 161005: gap of unknown length
* 161006 169238: contig of 8233 bp in length

Query Match 6.3%; Score 48.2; DB 2; Length 196650;

Best Local Similarity 42.9%; Pred. No. 7.4;

Matches 239; Conservative 0; Mismatches 318; Indels 0; Gaps 0;

QY 53 TTTTGTGATTCCTGGCCGTCGATTTGCGCGCTTGCGCATTCACATCACTATGATA 112
Db 98908 TTGTGTTTTTCTGCTGTTGTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 98967
QY 113 AATCACCAGAAATCAACAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 172
Db 98968 TTGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 99027
QY 173 AAACATAGATCCAAATGAAGTACGTATGATGATGATGATGATGATGATGATGATGATGATG 232
Db 99028 CTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 99087
QY 233 TTGTGATTTCAAGGTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 292
Db 99088 CTGCTGTTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTG 99147
QY 293 TGAACGTCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 352
Db 99148 TTGTTACTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 99207
QY 353 TGAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 412
Db 99208 TTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTG 99267
QY 413 ATCTTCATCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 472
Db 99268 CTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTG 99327
QY 473 TGAATATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 532
Db 99328 TTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTG 99387
QY 533 ATGTTGTCACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 592
Db 99388 TTGTTACTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 99447
QY 593 ATGTAATGACCGCTAAT 609
Db 99448 TTGTTGCTGTTGCTGTTG 99464

40059	40758:	gap of unknown length
40759	42636:	contig of 1878 bp in length
42637	42736:	gap of unknown length
42737	44174:	contig of 1438 bp in length
44175	44274:	gap of unknown length
44275	46152:	contig of 1878 bp in length
46153	46252:	gap of unknown length
46253	48003:	contig of 1751 bp in length
48004	48103:	gap of unknown length
48104	50040:	contig of 1937 bp in length
50041	50140:	gap of unknown length
50141	51569:	contig of 1429 bp in length
51570	51669:	gap of unknown length
51670	54119:	contig of 2450 bp in length
54120	54219:	gap of unknown length
54220	56090:	contig of 1871 bp in length
56091	56190:	gap of unknown length
56191	57453:	contig of 1263 bp in length
57454	57553:	gap of unknown length
57554	59060:	contig of 1507 bp in length
59061	59161:	gap of unknown length
59161	60625:	contig of 1465 bp in length
60626	60725:	gap of unknown length
60726	62447:	contig of 1722 bp in length
62448	62547:	gap of unknown length
62548	65509:	contig of 2962 bp in length
65509	65609:	gap of unknown length
65610	68150:	contig of 2541 bp in length
68151	68250:	gap of unknown length
68251	71440:	contig of 3190 bp in length
71441	71540:	gap of unknown length
71541	73743:	contig of 2203 bp in length
73744	73843:	gap of unknown length
73844	76497:	contig of 2654 bp in length
76498	76597:	gap of unknown length
76598	78364:	contig of 1767 bp in length
78365	78464:	gap of unknown length
78465	82749:	contig of 4285 bp in length
82750	82849:	gap of unknown length
82850	87186:	contig of 4337 bp in length
87187	87286:	gap of unknown length
87287	89122:	contig of 2626 bp in length
89133	90012:	gap of unknown length
90013	92515:	contig of 2503 bp in length
92516	92615:	gap of unknown length
92616	96753:	contig of 4140 bp in length
96756	96855:	gap of unknown length
96856	100641:	contig of 3766 bp in length
100642	100741:	gap of unknown length
100742	104699:	contig of 3958 bp in length
104700	104799:	gap of unknown length
104800	108873:	contig of 4074 bp in length

```
/organism="Rattus norvegicus"  
/db_xref="taxon:10116"
```

Query Match	6.3%;	Score 47.6;	DB 2;	Length 108873;
Best Local Similarity	35.08;	Pred. No. 10;		
Matches 110;	Conservative 0;	Mismatches 204;	Indels 0;	Gaps 0

QY	446	TTCACGATTTTGTGGTGCGCTTCGCCCTTGAATTTCGATGAAGTAACATAACAATGA	505
Db	23056	TGCAGTAGTGTGATGAACACTACTGTCATCGATAGNNNNNNNNNNNNNNNNNNNNNNNN	23115
QY	506	CATCTTTTGAAGACACAAATTCGTAATGTTTCAACCATATTGGTGTCTTTCATCT	565
Db	23116	NN	23175
QY	566	TGCATCCAATTTTGGCGCTTATTCGATGATTTGACCGCATTTTCCAGAACAACGTCAC	625
Db	23176	NNNNNNNNNNNNNNCCGCTGATCGATGGTAGTGGTTTCATCAGTTCAAAGATCAATGTA	23235

QY	626	GTAGGAATATGACCAAATGATTTGGCACCAGCATTTAAACGTAATTTGGAAAAAATTAAAC	685
Db	23236	CTTAATGAACTAACCTACACTTGAATAGCGCTTCCAAAGTGAAATTAATTTATTACAG	23295
QY	686	CAATGACATCATTTTTCCAACTGACAAATCTCATTTTCACAGCAATTAATAATTTT	745
Db	23296	GGTTATTAATATTTTTTGACCAAAAACATTTTTTGTGAAATTAATAAATGTAATAAT	23355
QY	746	TTATTTTATTTCT	759
Db	23356	AGATTTCTTTCAC	23369

RESULT 9	AC104634/c	AC104634	115758 bp	DNA	linear	PRI 10-AUG-2002
LOCUS						
DEFINITION		Homo sapiens	BAC clone RP11-753L24	from 2,	complete sequence.	
ACCESSION		AC104634				
VERSION		AC104634.5	GI:22002223			
KEYWORDS		HTG.				
SOURCE		human.				
ORGANISM		Homo sapiens				

REFERENCE
1 (bases 1 to 115758)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792

REFERENCE	2 (bases 1 to 115758)
AUTHORS	Du, H., Kozlowicz, A. and Spalding, L.
TITLE	The sequence of Homo sapiens BAC clone RP11-793L24

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (15-DEC-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 115758)

TITLE Waterston, R.H.
AUTHORS Direct Submission
JOURNAL Submitted (21-FEB-2002) Genome Sequencing Center, Washington

UNIVERSITY SCHOOL OF MEDICINE, 4444 FOREST PARK PARKWAY, ST. LOUIS,
MO 63108, USA
5 (bases 1 to 115758)
Unpublished R N
REFERENCE

AVINON
TITLE
JOURNAL
WATERSTON, R.H.
Direct Submission
Submitted (29-JUL-2002) Genome Sequencing Center, Washington
University, School of Medicine 4444 Forest Park Parkway St Louis

REFERENCE
6 (bases 1 to 115758)
MO 63108, USA
UNIVERSITY SCHOOL OF MEDICINE, 444 FOREST PARK DRIVE, D. BOULDER

AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2002) **Department of Genetics, Washington**

On Jul 29, 2002 this sequence version replaced gi:18855170.
Genome Center
University of California, San Francisco, California, USA
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: satishkharan.wustl@wustl.edu

CONTACT: sadpenn@wisc.edu
Summary Statistics
Center project name: H_NH0793L24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPcI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Feinguen, E., Tatenno, M., Catanesi, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-56215, 2000 bp overlap; the clone sequenced to the right is RP11-690121, 2000 bp overlap. Actual start of this clone is at base position 131499 of RP11-56215; actual end is at base position 65545 of RP11-690121.

Simple sequence repeats exists between 20489 and 21499.
Polymorphisms exist between AC017104, AC013476 and AC104634.

FEATURES

source
location/Qualifiers
1..115758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-793124"
/clone_lib="RPcI-11"
1..280
/rpt_family="Alu"
289..320
/rpt_family="(TAAA)n"
335..634
/rpt_family="Alu"
1043..1348
/rpt_family="Alu"
1356..1489
/rpt_family="Alu"
2044..2092
/rpt_family="(TAA)n"
2094..2424
/rpt_family="Alu"
2427..2605
/rpt_family="Alu"
2609..2676
/rpt_family="Alu"
2759..2790
/rpt_family="MaLR"
2791..3080
/rpt_family="AT-rich"
3083..3378
/rpt_family="Alu"
3441..3564
/rpt_family="L1"
3602..3643
/rpt_family="Alu"
3645..3838
/rpt_family="Alu"
3839..4051
/rpt_family="L1"
4300..4589
/rpt_family="Alu"
4686..4846
repeat_region

/rpt_family="L1"
5660..5955
/rpt_family="Alu"
6048..6347
/rpt_family="Alu"
6383..6447
/rpt_family="MaLR"
6448..6737
/rpt_family="Alu"
6743..6778
/rpt_family="(TTA)n"
6779..6807
/rpt_family="MaLR"
6808..6843
/rpt_family="ERV1"
6844..7149
/rpt_family="Alu"
7150..7382
/rpt_family="ERV1"
7383..7681
/rpt_family="Alu"
7682..7913
/rpt_family="ERV1"
7914..8310
/rpt_family="MaLR"
8335..8625
/rpt_family="Alu"
8820..9199
/rpt_family="MaLR"
9205..9366
/rpt_family="MER103"
9371..9686
/rpt_family="Alu"
9703..9875
/rpt_family="MaLR"
11178..11536
/rpt_family="ERV1"
11839..11880
/rpt_family="ERV1"
11881..12178
/rpt_family="Alu"
12179..12419
/rpt_family="ERV1"
12420..12722
/rpt_family="Alu"
12723..12876
/rpt_family="ERV1"
12877..13173
/rpt_family="Alu"
13174..13641
/rpt_family="ERV1"
14200..14389
/rpt_family="L1"
14403..14429
/rpt_family="(T)n"
14466..14776
/rpt_family="Alu"
14825..14878
/rpt_family="(TC)n"
16155..16213
/rpt_family="MIR"
16214..16523
/rpt_family="Alu"
16625..16848
/rpt_family="MaLR"
16854..17157
/rpt_family="Alu"
17165..17277
/rpt_family="MaLR"
repeat_region

Query Match 6.2%; Score 47.4; DB 9; Length 115758;
Best Local Similarity 45.7%; Pred. No. 11;
Matches 165; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

Oy	134	AAGCATTTGATGCATCCATCGCTGATTGGAAACATCCGAACAACATPACTCAATGAAGA	193
Db	21431	ATGATGTGTCAGCAGTGATGTGTGTGTGATGATGATGATGATGATGATGATGATGATG	21372
Oy	194	TACCGATCATGCCGATAAATTCGAACCTCATGTTGTGATTTGTGGATTTCAAAGTGAAT	253
Db	21371	ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	21312
Oy	254	TAGCATCCGAACATTGAGCGTCGAGATTGGAACAACAATGAAACGTCAAGTGTGCTA	313
Db	21311	ATGACAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	21252
Oy	314	ATGCAAGGTGAAGAGAGGTAATTTGTAAGCTCATTTGTGATGCGTGTCAAGATGATA	373
Db	21251	GTCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	21192
Oy	374	TGCTTCGATGAATGATGATTTAGCATCAAAATGGCGATCTTCATCCAACACTCATG	433
Db	21191	CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	21132
Oy	434	TCATTTCCGATATTCACATTTTTGTTGCTTGTCCCTTGAAATTTCTGATGAAGTA	493
Db	21131	TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	21072
Oy	494	A 494	
Db	21071	A 21071	
RESULT 10			
AC105625/c		147671 bp DNA linear HTG 13-JUN-2002	
LOCUS	Rattus norvegicus clone chf230-117017, *** SEQUENCING IN PROGRESS		
DEFINITION	***, 70 unordered pieces.		
ACCESSION	AC105625		
VERSION	AC105625.2 GI:21736522		
KEYWORDS	HTG; HTGS PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus Euharjota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 147671)		
REFERENCE	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshirooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbiera,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunyah,J., Buruch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Daylan,M.L., Davis,C., Davy-Carroil,L., Dederich,D.A., DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gotrell,D.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kuresht,A., Landuy,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgeid,H., Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Mathewswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Mosser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okumonu,G., Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,		

TITLE	JOURNAL
AUTHORS	AUTHORS
TITLE	TITLE
JOURNAL	JOURNAL
REFERENCE	REFERENCE
AUTHORS	AUTHORS
TITLE	TITLE
JOURNAL	JOURNAL
COMMENT	COMMENT
<p>Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojs,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtai,N., Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D., Weinstock,G., and Gibbs,R.</p> <p>Direct Submission Unpublished 2 (bases 1 to 147671) Worley,K.C. Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 147671) Worley,K.C. Direct Submission Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 12, 2002 this sequence version replaced gi:18092848.</p> <p>----- Genome Center ----- Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information ----- Center project name: GMS Center clone name: CH230-117017 ----- Summary Statistics ----- Sequencing vector: Plasmid: Chemistry: Dye-terminator Big Dye: 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 80937 bases at least Q40 Consensus quality: 85186 bases at least Q30 Consensus quality: 88448 bases at least Q20</p> <p>----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html) * NOTE: This is a "working draft" sequence. It currently * consists of 70 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.</p> <p>1 1166: contig of 1166 bp in length * 1167 1286: gap of unknown length * 1267 2389: contig of 1123 bp in length * 2390 2489: gap of unknown length * 2490 3664: contig of 1175 bp in length * 3665 3765: gap of unknown length * 3765 4870: contig of 1106 bp in length * 4871 4970: gap of unknown length * 4971 6060: contig of 1090 bp in length * 6061 6161: gap of unknown length * 6161 7186: contig of 1025 bp in length * 7186 7285: gap of unknown length * 7285 8523: gap of unknown length * 8524 8623: gap of unknown length * 8624 10363: contig of 1740 bp in length * 10364 10463: gap of unknown length * 10464 11859: contig of 1396 bp in length * 11860 11959: gap of unknown length * 11960 13257: contig of 1298 bp in length * 13258 13357: gap of unknown length * 13358 14373: contig of 1016 bp in length</p>	

[illegible]

gene	/evidence=experimental /rpl_type=Inverted complement(2443. .2631) /gene="MSV001" complement(2443. .2631) /gene="MSV001" /codon_start=1 /product="ORF MSV001 62 amino acid repeat gene family protein" /protein_id="AAC97866.1" /db_xref="GI:4049906" /translation="MFLLISQNNINNSKISISNTQSMFLLYLSQNNINNSKISINT QSMFLLYLSQNNINNSKI"
CDS	complement(6051. .6320) /gene="MSV002" complement(6051. .6320) /gene="MSV002" /codon_start=1 /product="ORF MSV002 89 amino acid repeat gene family protein" /protein_id="AAC97867.1" /db_xref="GI:4049907" /translation="MAEKRRYKVKLEPVDYKNSLSTIEALQNEIQKSIDQSNILKQ NLSEIDIMNEVEFESQVETLIMLNKNIYISTEEKIDYNSLS"
gene	complement(6347. .6814) /gene="MSV004" complement(6347. .6814) /gene="MSV004" /codon_start=1 /product="ORF MSV004 155 amino acid repeat gene family protein" /protein_id="AAC97868.1" /db_xref="GI:4049908" /translation="MYSCMKMEDECYVLCKNCNWSYQYTDGICINDPPINKLDD CKRCKRINPVDHGVCGEYANKIVCGNNTSTCTIILINDQKIEGVSSSDIFIT SKRFYLIINIVVLIISNNKILQIINNIDKRYEVFNKYLLENFVHILVQOIDEI"
CDS	complement(6574. .6756) /gene="MSV003" complement(6574. .6756) /gene="MSV003" /codon_start=1 /product="ORF MSV003 60 amino acid repeat gene family protein" /protein_id="AAC97869.1" /db_xref="GI:4049909" /translation="MYHLINIQMWALIVIFLLILNLIIYVINTKLLIFMKRYAVNM LTKLYVMVLIQNLHYI"
gene	complement(7431. .8171) /gene="MSV005" complement(7431. .8171) /gene="MSV005" /codon_start=1 /product="ORF MSV005 hypothetical protein" /protein_id="AAC97860.1" /db_xref="GI:4049900" /translation="MKMKILFDQVNTKLIIEYIKSLDQGNLKKKLNIDDKLSIKIK LTCVAPNISECPISITIKFYLITLSEIILILINLYKQNLNRAKHNMYKKYC KCECILTQNPQPEICEICYKKNKVPFGCHGCHDITVEDELNSCKNCNPFINKSYATIG VLMKRIKIKICLIGNDIYISNICSECKYVLFKRLFTKRCIQCNMFIDSELTATKCC HGCYRYVCWGWGYWYLYPLFDYNG"
gene	8831. .9058 /gene="MSV006" 8831. .9058 /gene="MSV006" /codon_start=1 /product="ORF MSV006 hypothetical protein" /protein_id="AAC97608.1" /db_xref="GI:4049648" /translation="MEESLPSDMCNIILRLINDSKNAKFLIIQELHFSINDISNVHSY IPEFLIPTSNLIIINSKNUKISIOSLO"
CDS	complement(9825. .11327) /gene="MSV008" complement(9825. .11327) /gene="MSV008"

gene
/gene="MSV008"
/product="ORF MSV008 leucine rich repeat gene family
protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
SW:P28854"
/protein_id="AAC97859.1"
/db_xref="GI:4049899"
/translation="MNNINLYCYNNILKYLDNDNDKNNIGIYELFKYYNCLKWPYV
KFEFYNNLEYIKNLNDCSDNITDFEESLINLEILDISYNNNSIYKCLSLKLNKL
NCKEKKIDFELESLINLEILDISYNNNSIYKCLSLKLNCKEKKIDFELE
SLINLEILDYSNNKSNITECTPISELKCIICKIDREFEESLINLEILDYSNE
YNSIDCKIPVSLKLNCEFCYIDFEKLEPLNNIKLNTSYNNNSISCKKLPVSL
IKLNCGCYIDFEKLEPLNLOKLNISNSYSISCKPISLIKRSEYCDITNPK
FERLINIETETIYMNSSRSISCKLPSPLEINLEIWDIKINIEKLEIKLNLSL
SSSSNNISCIPTSLIYDLISCHIKRFEKLEPLNLEIKLNISDNGSSISCKLP
SLIELYCKCKNNKFNFIQPLINIQKIDISVNDYSSNIILNHFYPPFIN"
10282..10476
/gene="MSV007"
/gene="MSV007"
10282..10476
/gene="MSV007"
/codon_start=1
/product="ORF MSV007 hypothetical protein"
/protein_id="AAC97609.1"
/db_xref="GI:4049649"
/translation="MYIVNSKLINRSKNLKFVISOYSDNLINDIGNLHFDIFEYL
ELIFNCKIKNSKRLKSYI"
complement(11374..12138)
/gene="MSV009"
complement(11374..12138)
/gene="MSV009"
/codon_start=1
/product="ORF MSV009 leucine rich repeat gene family
protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
SW:P28854"
/protein_id="AAC97858.1"
/db_xref="GI:4049898"
/translation="MNNLTYYRKEKLKLNLDNLIKIHKLAAANKLYKRYNSILNPLHE
NIKYNNPRHYAKLQCGKIPNNLEKILNCELTICGASKIESKLPISLKIYFNNK
ITDFEELINLEKLTPIIKNKSNISCKPISTIIFCCNCKEIDFEKLEPLNLYK
KIYISGKNNSNSHCNDEPIKTELEPCSCNLTDFKFLKNLYNKLKLEPTGNNNSIN
CELPRKLKFCVINCNIITDFENFKSSNNLRITITSN"
complement(12364..14199)
/gene="MSV010"
complement(12364..14199)
/gene="MSV010"
/codon_start=1
/product="ORF MSV010 leucine rich repeat gene family
protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
SW:P28854"
/protein_id="AAC97857.1"
/db_xref="GI:4049897"
/translation="MNNINLYCYENLIKLYKSDIHNLASICEFQYNYNQLNCVNH
LYKNNYKRYIKRNFSSDRFGLTITDFEELNNLEILDISNNEKLNISCKLPKI
KKFICVNDIDDFEKLPELNLMEVLDISYNNNSIYKSLQSKDCEKCNITPT
FLKLNINNKLDISENTSYNSISCKLPRTIKFPCDRCIDITDFEELINLEILDIS
NIIINISNYKSLISIKKEICARCAITDFEELNLMEILDVSYNNKLNISCELP
SLIKKIYCNCEPKNNTLKLKLNIFINISNPKITDPKLEPLNLEILDYSNNKSN
ISCKLSTKTIKFCISRCATIDFEKLEPLNLEILDVSYNTEANISCFKISIKLEIN
CSDKLTIDDFEFLERLINLOKLNISYFSSNISCELTSTVLNENISCTCKINNFTIE
KLYNKLKIDINNNPNSISCKLTALIELEDCYTICNTDFKLEPLINLOKLNISL
DNSIKIETPLSLIENCMYCIISDFKLEPLISKLILINCKNGSSNIYCNLSPILE
TIEYKHHIDPEYDINNSLHKLVYDFTSDSDSDSDSDSGIYD"
14293..15810
/gene="MSV011"
14293..15810
/gene="MSV011"
/codon_start=1
/product="ORF MSV011 leucine rich repeat gene family
protein, similar to Amsacta moorei entomopoxvirus Q3 ORF
SW:P28854"
/protein_id="AAC97610.1"
/db_xref="GI:4049650"
/translation="MNNINLYCYENLLKYLDCKRHKFLAKYSKIFQYNNYQNLNPIHY

CDs

OTSTVINKLENNKSFLLAMTLEPLIDQYNSDFVFIKLQNSKYYNHFSLDDIDGSI
BDRLDVYFI"
6934..7811
/note="Orf17: similar to ACORF141 homolog"
/codon_start=1
/product="le-0"
/protein_id="AAK96301.1"
/db_xref="GI:15384390"
/translation="MSGTTLKRLIYDDISDSQAKLFERYNSEMPAPASQOMNTAVDE
IDVEYIKCFEKLKNMYSDDVTNARQYVVKLAAPLIYDEKKQYKNNLKHSLYVYK
ESESIVILDEQOCHHTLPIITLOOLIKTICYLIMNSDEENVYQKFTLEPLYLEK
ILKEIYQVCKCKLTKLOALOANTLITQSDASCHHAIINFSOVLIVPLENPLYEON
ICRDFNEBRHTRKPECCGYKICNLCLYANLMKYSTVPTCPVCATSRKSSSVSFQVY
YTAQTTDNT"

CDs

7828..9234
/note="Orf8"
/codon_start=1
/product="p49"
/protein_id="AAK96264.1"
/db_xref="GI:15384353"
/translation="MNLDEKNVGLERNNNYKYLFLASYENLADTGILSTSKKPIREYL
YNNENINDASLIGLDYLDLIGLNRYLIDDRVNMFKYIKPQFPEVCTKAVELLKED
OBVYIKPPTVPATNEFPKVPNSSEKFLIYVNSVYIQRKNFVNDKNDKNCILQNGTGVV
EDQAVDMCGVRCVPEKIELESSPEPRILVLDAAHRHATNNISPDSENILKPKF
YKGLMPEFRNTRKIINSKFFTRKPNHLNEFKQEDRTKSAYVFTIQRYIYDAAYDPD
DLILDNLNHSMTYSVYKFTKEMEDGEPGVYSEIYDRAVDKYOXLSIKIDETIM
FPTLLNDSPLYEIRPDLIQIKGTLMARVYKHKLFALIANNSLFSGTTLLEPRKLI
PFRQYQPPRLIADDERYVVDKQKLYLKYTPANTIPAYLLIRGYESSEIKTLRLDK
PMVONTLKLIIAAPPSK"
9245..9490
/note="Orf9"
/codon_start=1
/product="odv-e18"
/protein_id="AAK96265.1"
/db_xref="GI:15384354"
/translation="HDDLRGTTTGAGRPNNMLNPSMLTILALVILLLMLFOS
SEPSGKADTNAFAQNPFLNATMRNPPVNPQRTIM"
9505..10359
/note="Orf10"
/codon_start=1
/product="odv-ec27"
/protein_id="AAK96266.1"
/db_xref="GI:15384355"
/translation="MKKRCQSNKIRVTEIINADEKLHKDYDLADFNAKNLSLESY
DNLQIKMILLAKYAMALNMLLELPOLPLATFRDKNAIKEIVSIYFASLGVRHVRVPMIN
HNSMEKIVLENRNASIJGELPFCOHNDGVICYIDRPSILMSKDPDLPVDMNIN
MKKENQYMIKATCFRCAPKRHSRREPRPEINLTEDVQYQWTLFIENHAYLHYX
ILKNKGYVDYSLSLDHTLFSKSKSPRTILNMFKNILLSKRFSLIEDSDISYTKNTKNNK
LGILTYTD"

CDs

10404..10682
/note="Orf11: similar to ACORF145 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAK96267.1"
/db_xref="GI:15384356"
/translation="MMLLALFELIYKLLVYHKMONLQYDMHHRKLCBPAGYGNLADPEP
DQNAVYCEPKIKFYCPKRNQYFNIDPAQCQPSDLETGCIQYNNRLLL"
complement(10709..11314)
/note="Orf12: similar to ACORF146 homolog"
/codon_start=1
/product="ep23"
/protein_id="AAK96268.1"
/db_xref="GI:15384357"
/translation="MNINLYOPLSIHIDLIVFTIYDAFSVTVYVNAKLDVINDQNGCS
PRTVYITKLVSCHESGRNYSMNKSTTKSGAIKDVYFITMKKLPYVARSLSLEDKCSRS
RVLVLADYNNSTELIMHVLVYRRSRNNAKCYERYDALETTDSAGNELYSKEHYILK
GNVPAAPFASLAKAATSCDMINITHPHVYIINSKRVITITH"
11356..11341
/note="Orf13: immediate early 1"
/codon_start=1
/product="le-1"
/protein_id="AAK96269.1"
/db_xref="GI:15384358"

CDs

[illegible]

QY	759	TC	760
Db	19186	TC	19187

VERSION
KEYWORDS
SOURCE
ORGANISM

AC009630.6 GI:22450494
HTG:HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.

REFERENCE

Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,

LOCUS	1141 bp	DNA	Linear	PAT 28-FEB-2001
AX083744				
DEFINITION	Sequence 22 from Patent WO0111061.			
ACCESSION	AX083744			
VERSION	AX083744.1	GI:13185472		

RESULT 16
10000000

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 236542)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department

COMMENT

On Aug 23, 2002 this sequence replaced gi:21729491.

Center: Baylor College of Medicine

```
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GEXP
Center clone name: CH230-169E3
----- Summary Statistics -----
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 212949 bases at least Q40
Consensus quality: 214002 bases at least Q30
Consensus quality: 214801 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_direct_data.html).
* NOTE: This is a working draft sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1076: contig of 1076 bp in length
* 1077 1176: gap of unknown length
* 1177 2315: contig of 1139 bp in length
* 2316 2415: gap of unknown length
* 2416 3451: contig of 1036 bp in length
* 3452 3552: gap of unknown length
* 3552 4591: contig of 1040 bp in length
* 4592 4691: gap of unknown length
* 4692 5858: contig of 1167 bp in length
* 5859 5958: gap of unknown length
* 5959 6974: contig of 1016 bp in length
* 6975 7075: gap of unknown length
* 7075 8437: contig of 1363 bp in length
* 8438 8537: gap of unknown length
* 8538 9689: contig of 1152 bp in length
* 9690 9789: gap of unknown length
* 9790 10848: contig of 1059 bp in length
* 10849 10949: gap of unknown length
* 10949 12163: contig of 1215 bp in length
* 12164 12263: gap of unknown length
* 12264 13582: contig of 1319 bp in length
* 13583 13682: gap of unknown length
* 13683 14793: contig of 1111 bp in length
* 14794 14893: gap of unknown length
* 14893 16268: contig of 1375 bp in length
* 16269 16368: gap of unknown length
* 16369 19601: contig of 3233 bp in length
* 19602 19702: gap of unknown length
* 19702 26083: contig of 6382 bp in length
* 26084 26183: gap of unknown length
* 26184 36956: contig of 10773 bp in length
* 36957 37056: gap of unknown length
* 37057 48029: contig of 10973 bp in length
* 48030 48129: gap of unknown length
* 48130 61614: contig of 13485 bp in length
* 61615 61715: gap of unknown length
* 61715 75946: contig of 14232 bp in length
* 75947 76046: gap of unknown length
* 76047 93495: contig of 17449 bp in length
* 93496 93595: gap of unknown length
* 93596 117393: contig of 23798 bp in length
* 117394 117493: gap of unknown length
* 117494 142433: contig of 24940 bp in length
* 142434 142533: gap of unknown length
* 142534 178061: contig of 35528 bp in length
* 178062 178161: gap of unknown length
* 178162 236542: contig of 58381 bp in length.
*
* Location/Qualifiers
* 1..236542
```

```
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-169E3"
BASE COUNT 65943 a 45570 c 46993 g 64960 t 13076 others
ORIGIN
Query Match 6.0%; Score 45.8; DB 2: Length 236542;
Best Local Similarity 44.1%; Pred. No. 21;
Matches 237; Conservative 0; Mismatches 297; Indels 3; Gaps 1;
Oy 63 TGTGCGCGCATTTGTCGCCGTTTCGGCTGATCAATTCATGATTAATAATCACCAG 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156421 TGTTCCTGCTGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 156480
Oy 123 AGAATCAACAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATG 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156481 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156540
Oy 183 TCAATGAAGTACCTGATCAATGATCAATGATCAATGATCAATGATCAATGATCAAT 242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156541 TGTTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 156600
Oy 243 CAAGTGAATTAGC--CATGCAAAACATTGAGGCTGAGAGATTGAAACAATGAAACG 299
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156601 AGCTGTTGCTGCTGCTGTTGTTGTTGATGATGATGATGATGATGATGATGAT 156660
Oy 300 TCAAGTATGCTATGCTCAAGGATGAAAGAGGATATTTAAGTCAATTTGTTGATCGG 359
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156661 TGTACCTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 156720
Oy 360 TGTTCACGATGATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 419
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156721 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156780
Oy 420 TCAACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 479
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156781 TGTTCCTGTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 156840
Oy 480 TTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 539
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156841 TGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156900
Oy 540 CAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 596
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156901 TGTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 17
AC115666 169163 bp DNA linear HTG 18-JUN-2002
LOCUS AC115666
DEFINITION Rattus norvegicus clone CH230-261G15, *** SEQUENCING IN PROGRESS
ACCESSION AC115666
VERSION AC115666.3 GI:21745940
KEYWORDS HTG: HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 169163)
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaral-tunget,H.C., Are,J.R., Aylee,M., Banks,T.,
Barbarta,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Buck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dalhorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
```

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabrill,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorbett,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgeson,A., Hognes,M., Holloway,C., Hollins,B., Homti,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Jouda,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newsum,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S., Oguh,M., Okunnu,G., Oranuyne,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., RojudoKan,T., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoozhari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vaquez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,C., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 169163)
Worley,K.C.

Direct Submission
Submitted (22-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 169163)
Worley,K.C.

Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:21240080.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTRP
Center clone name: CH230-261G15
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 127101 bases at least Q40
Consensus quality: 131944 bases at least Q30
Consensus quality: 136029 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1195: contig of 1195 bp in length
* 1196 1295: gap of unknown length
* 1296 2301: contig of 1006 bp in length
* 2302 2401: gap of unknown length

2402 3597: contig of 1196 bp in length
* 3598 3697: gap of unknown length
* 3698 4705: contig of 1008 bp in length
* 4706 4805: gap of unknown length
* 4806 6053: contig of 1248 bp in length
* 6054 6153: gap of unknown length
* 6154 7709: contig of 1536 bp in length
* 7710 7809: gap of unknown length
* 7810 9075: contig of 1266 bp in length
* 9076 9175: gap of unknown length
* 9176 10387: contig of 1212 bp in length
* 10388 10487: gap of unknown length
* 10488 12171: contig of 1684 bp in length
* 12172 12271: gap of unknown length
* 12272 13923: contig of 1652 bp in length
* 13924 14023: gap of unknown length
* 14024 15253: contig of 1230 bp in length
* 15254 15353: gap of unknown length
* 15354 17056: contig of 1703 bp in length
* 17057 17156: gap of unknown length
* 17157 18630: contig of 1474 bp in length
* 18631 18730: gap of unknown length
* 18731 20564: contig of 1834 bp in length
* 20565 20664: gap of unknown length
* 20665 22920: contig of 2256 bp in length
* 22921 23020: gap of unknown length
* 23021 24042: contig of 1022 bp in length
* 24043 24142: gap of unknown length
* 24143 25407: contig of 1265 bp in length
* 25408 25507: gap of unknown length
* 25508 27326: contig of 1819 bp in length
* 27327 27426: gap of unknown length
* 27427 28825: contig of 1399 bp in length
* 28826 30295: gap of unknown length
* 30296 30395: contig of 1370 bp in length
* 30396 32901: gap of unknown length
* 32902 33001: gap of unknown length
* 33002 35383: contig of 2382 bp in length
* 35384 35483: gap of unknown length
* 35484 38276: contig of 2793 bp in length
* 38277 38376: gap of unknown length
* 38377 39868: contig of 1492 bp in length
* 39869 39968: gap of unknown length
* 39969 41969: contig of 2001 bp in length
* 41970 42069: gap of unknown length
* 42070 44382: contig of 2313 bp in length
* 44383 44482: gap of unknown length
* 44483 47592: contig of 3110 bp in length
* 47593 47692: gap of unknown length
* 47693 52151: contig of 4459 bp in length
* 52152 52251: gap of unknown length
* 52252 56760: contig of 4509 bp in length
* 56761 56860: gap of unknown length
* 56861 61177: contig of 4317 bp in length
* 61178 61277: gap of unknown length
* 61278 66148: contig of 4871 bp in length
* 66149 66248: gap of unknown length
* 66249 71886: contig of 5638 bp in length
* 71887 71986: gap of unknown length
* 71987 78079: contig of 6093 bp in length
* 78080 78179: gap of unknown length
* 78180 81154: contig of 2975 bp in length
* 81155 81254: gap of unknown length
* 81255 88890: contig of 7636 bp in length
* 88891 88990: gap of unknown length
* 88991 97276: contig of 8266 bp in length
* 97277 97376: gap of unknown length
* 97377 107376: contig of 10000 bp in length
* 107377 107476: gap of unknown length
* 107477 116754: contig of 9278 bp in length
* 116755 116854: gap of unknown length
* 116855 127329: contig of 10475 bp in length

```

* 127330 127429: gap of unknown length
* 127430 136746: contig of 9317 bp in length
* 136747 136846: gap of unknown length
* 136847 153947: contig of 17101 bp in length
* 153948 154047: gap of unknown length
* 154048 169163: contig of 15116 bp in length.
FEATURES
    source
        1..169163
        /organism="Rattus norvegicus"
        /db_xref="taxon:10116"
        /clone="CH230-261G15"
BASE COUNT  41981 a 37857 c 38661 g 43240 t 7424 others
ORIGIN
Query Match      5 9%: Score 45.2; DB 2; Length 169163;
Best Local Similarity 46.7%: Pred. No. 28;
Matches 143: Conservative 0; Mismatches 163; Indels 0; Gaps 0;
OY 304 GGTGATGCTAATGTCMAAGAGAGGATGATTAAGCTCAATTTGTTGATCGGTGT 363
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34249 GTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34308
OY 364 CACGATGATATCGTCTCGATGGAATATGATTAGCATACAAATGGGATCTTCATCA 423
DB 34309 GTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34368
OY 424 ACCACTCAGTCATTTTCCGATATTCAGATTTTGTGTTGCTTGCTTGAATTTCT 483
DB 34369 GTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34428
OY 484 GATGAAGGTAACATTAACATGACATCTTTTGAAGTACGACAAATCGATATTTGCAAC 543
DB 34429 GCTGTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34488
OY 544 CATATGCTGCTTTCAATCTTGATCCAAATTTTGGCGTTTATCTGATTAATGACC 603
DB 34489 GTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34548
OY 604 GCTATT 609
DB 34549 GTTGT 34554

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Telamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemdek,L., Zimmer,A., and Zody,M.
Direct Submission
Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 175119)
TITLE
JOURNAL
REFERENCE
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,Y., Murphy,T., Naylor,J., Nguyen,C.,
Nicoli,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S.,
Schnupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Telamas,J., Testaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zemdek,L., Zimmer,A., and Zody,M.
Direct Submission
Submitted (02-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 24, 2002 this sequence version replaced gi:21104967.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web Site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19879
Center clone name: 424.L.10
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 169912 bases at least Q40
Consensus quality: 171891 bases at least Q30
Consensus quality: 172998 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 173819; sum-of-coverage
Quality coverage: 8.4 in Q20 bases; agarose-fp
Quality coverage: 8.5 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 736: contig of 736 bp in length
* 737 836: gap of 100 bp

```


[illegible]

SOURCE	Homo sapiens.	AUTHORS	Worley, K.C.
ORGANISM	Homo sapiens	TITLE	Direct Submission
REFERENCE	Mammalia: Eutheria: Primates; Catarrhini: Hominoidea: Homo.	JOURNAL	Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
AUTHORS	1 (bases 1 to 172853)	COMMENT	On Jul 28, 2001 this sequence version replaced gi:14861680. INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu
	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstons, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barbieri, J., Benton, J., Blincoe, K., Blomquist, K., Bonin, D., Bouck, J., Bowe, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burkh, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, R.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douhaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsli, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, U., Korvah, J., Kovar, C., Kravtsov, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtenarge, O., Lieu, C., Liu, J., Liu, W., Louissege, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marandel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, N., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenwo, S., Ogulu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Prims, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shoshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczka, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Zuchelapati, R., Weinstock, G. and Gibbs, R.		
TITLE	Direct Submission		
JOURNAL	Unpublished	FEATURES	QUALSTAT-REPORT.
REFERENCE	2 (bases 1 to 172853)	source	Location/Qualifiers
AUTHORS	Worley, K.C.		1..172853
TITLE	Direct Submission		/organism="Homo sapiens"
JOURNAL	Submitted (26-OCT-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		/db_xref="taxon:9606"
REFERENCE	3 (bases 1 to 172853)		/chromosome="12"
AUTHORS	Worley, K.C.		/clone="RP11-364C11"
TITLE	Direct Submission		1..2006
JOURNAL	Submitted (28-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		/note="Overlaps bases 98325..100330 of clone AC087885"
REFERENCE	4 (bases 1 to 172853)		/function="clone overlap"
AUTHORS	Worley, K.C.		complement(2..154)
TITLE	Direct Submission		/rpt_family="MIR"
JOURNAL	Submitted (28-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		complement(1151..1320)
REFERENCE	5 (bases 1 to 172853)		/rpt_family="MIR"
AUTHORS	Worley, K.C.		1623..1653
TITLE	Direct Submission		/rpt_family="(A)n"
JOURNAL	Submitted (27-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		1653..1832
REFERENCE	6 (bases 1 to 172853)		/rpt_family="MER117"
AUTHORS	Worley, K.C.		complement(1972..2046)
TITLE	Direct Submission		/rpt_family="MIR"
JOURNAL	Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		2146..2246
REFERENCE			/rpt_family="12"
AUTHORS	Worley, K.C.		complement(2971..3160)
TITLE	Direct Submission		/rpt_family="MIR"
JOURNAL	Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		3474..3525
REFERENCE			/rpt_family="MIR"
AUTHORS	Worley, K.C.		complement(4390..4760)
TITLE	Direct Submission		/rpt_family="MUT117"
JOURNAL	Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		complement(4883..5058)
REFERENCE			/rpt_family="MIR"

Worley, K.C.
Direct Submission
Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 28, 2001 this sequence version replaced gi:14861680.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

Location/Qualifiers

source

1..172853

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="12"

/clone="RP11-364C11"

1..2006

/note="Overlaps bases 98325..100330 of clone AC087885"

/function="clone overlap"

complement(2..154)

/rpt_family="MIR"

complement(1151..1320)

/rpt_family="MIR"

1623..1653

/rpt_family="(A)n"

1653..1832

/rpt_family="MER117"

complement(1972..2046)

/rpt_family="MIR"

2146..2246

/rpt_family="12"

complement(2971..3160)

/rpt_family="MIR"

3474..3525

/rpt_family="MIR"

complement(4390..4760)

/rpt_family="MUT117"

complement(4883..5058)

/rpt_family="MIR"

[illegible]

RESULT 24	
LOCUS	AC116979
DEFINITION	AC116979 107739 bp DNA linear HTG 22-MAY-2002
ACCESSION	AC116979
VERSION	AK4, *** SEQUENCING IN PROGRESS ***, In ordered pieces.
KEYWORDS	
SOURCE	AC116979.1 GI:20042934
ORGANISM	HTG: HTGS_PHASE2.
REFERENCE	Dictyostelium discoideum.
AUTHORS	Dictyostelium discoideum.
	Eukaryota: Mycetozoa; Dictyostellida: Dictyostelium.
	1 (bases 1 to 107739)
	Gloeckner,G., Eichinger,L., Szafrański,K., Pachebat,J., Dear,P.,
	Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
	Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
	Noegel,A.A.
TITLE	Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL	Unpublished
REMARK	The Dictyostelium Genome Sequencing Consortium
REFERENCE	2 (bases 1 to 107739)
AUTHORS	Baumgart,C.
TITLE	Direct Submission
JOURNAL	Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
	Biotechnology, Beutenbergstr. 11, Jena 0745, Germany
REFERENCE	3 (bases 1 to 107739)
AUTHORS	Baumgart,C.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAY-2002) Genome Analysis, Institute of Molecular
	Biotechnology, Beutenbergstr. 11, Jena 0745, Germany
COMMENT	CDS predictions from GenEdit may contain errors. Further Information
	is available from IMB Jena, Department of Genome Analysis
	(http://genome.imb-jena.de/dictyostelium/)
	and the University Cologne, Institute for Biochemistry I
	(http://www.uni-koeln.de/dictyostelium/project.shtml)
	Funding
	Agency : Deutsche Forschungsgemeinschaft (DFG).
	* NOTE: This is a 'working draft' sequence.
	* This sequence will be replaced
	* by the finished sequence as soon as it is available and
	* the accession number will be preserved.
FEATURES	Location/Qualifiers
source	1..107739
	/organism="Dictyostelium discoideum"
	/strain="AX4"
	/db_xref="taxon:44689"
	/chromosome="2"
	/map="6208312-6316049"
	join(246..316,462..1643)
	/note="ORF_ID:dd_00463"
	/pseudo
	/codon_start=1
	join(2659..2806,2951..3060,3360..3969,4108..4428,
	5384..6588)
	/note="ORF_ID:dd_01980"
	/pseudo
	/codon_start=1

Query Match 5.8% Score 44 DB 2 Length 107739;
 Best Local Similarity 45.6% Pred. No. 51;
 Matches 230; Conservative 0; Mismatches 270; Indels 4; Gaps 2;

QY 10 TCATTAACATCCAAAAACATATCTTACAAATGATGAATTTTGTGATTCCTCC 69
 Db 19725 TAATCAATCATGATAGACAAATCTTAAGAGAGTAATATATCAAGATTATCA 19784
 QY 70 GTGGCAT-TTGTCCCGCTTGGCGCTGATTCATTCACATATGATTAATTCACCGAAGAAAT 128
 Db 19785 TCACATCGTGTAGATATTTTCAAGCTTGAAGTGAAGTCTGAAAGTTATCATTTAAC 19844
 QY 129 CAACAAGCTATGATGATGCGATGCTGCTATTTGAACATCCGAAACATAGATCCAAAT 188
 Db 19845 AAATGATATATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 19904
 QY 189 GAAAGTACCTGATGATGCGGATTAATTCGAAGTGTGATTTGTGATTTCAAGG 248
 Db 19905 TGAAGAAAGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 19964
 QY 249 TGAATTAGCCATGCGAAACATTTGAGGCTCGAGATTGAAACAATGAACGTCAGGTGA 308
 Db 19965 TGAAGAAAGTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 20024
 QY 309 TGTATATGTCGAAGGTGAAGGATGATTTGAAGCTGATTTGATCGGTGATTCAGCA 368
 Db 20025 TTGCGAATTTGATCAACAGAGAAATATGATTTGATCAATATTTTGAAGAGTATTT 20084
 QY 369 TGATATGCTCTGATGGAATATGATTTGATGATCAATTTTGGTGATTTTCAACAC 425
 Db 20085 ATTCAAAATGATTAAGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTT 20144
 QY 426 CACATCATCTATTCGATATTCAGATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485
 Db 20145 CCTTCAATTCAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20204
 QY 486 TGAAGTAACTAACAATGATGATC 509
 Db 20205 TAAATTAATCTGAAGTATGCTTC 20228

RESULT 25
 AC096212/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-23N22, *** SEQUENCING IN PROGRESS
 AC096212 165337 bp DNA linear HTG 11-JUL-2002
 AC096212.4 GI:21723349
 VERSION HTG: HTGS PHASE1.
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 165337)
 Musny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alshrooks,S.L., Ameralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbatala,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,

Homs1,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratochic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,D., Liu,W., Louised,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogulu,M., Okwundu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Morley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

Unpublished
 Direct Submission
 2 (bases 1 to 165337)
 Morley,K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 165337)
 Morley,K.C.
 Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:17943903.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GERC
 Center clone name: CH230-23N22
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 96697 bases at least Q40
 Consensus quality: 103304 bases at least Q30
 Consensus quality: 107916 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 66 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 1019: contig of 1019 bp in length
 * 1020
 * 1119: gap of unknown length
 * 1120
 * 2182: contig of 1063 bp in length
 * 2183
 * 2282: gap of unknown length
 * 2283
 * 3490: contig of 1208 bp in length
 * 3491
 * 3590: gap of unknown length
 * 3591
 * 4592: contig of 1002 bp in length
 * 4593
 * 4692: gap of unknown length
 * 5977: contig of 1285 bp in length

```
* 5978 6077: gap of unknown length
* 6078 7186: contig of 1109 bp in length
* 7187 7286: gap of unknown length
* 7287 8413: contig of 1127 bp in length
* 8414 8513: gap of unknown length
* 8514 9682: contig of 1169 bp in length
* 9683 9782: gap of unknown length
* 9783 10917: contig of 1135 bp in length
* 10918 11017: gap of unknown length
* 11018 12051: contig of 1034 bp in length
* 12052 12151: gap of unknown length
* 12152 13374: contig of 1223 bp in length
* 13375 13474: gap of unknown length
* 13475 14588: contig of 1114 bp in length
* 14589 14688: gap of unknown length
* 14689 16062: contig of 1374 bp in length
* 16063 16162: gap of unknown length
* 16163 17280: contig of 1128 bp in length
* 17291 17390: gap of unknown length
* 17391 18879: contig of 1489 bp in length
* 18880 18979: gap of unknown length
* 18980 20557: contig of 1578 bp in length
* 20558 20657: gap of unknown length
* 20658 21994: contig of 1337 bp in length
* 21995 22094: gap of unknown length
* 22094 23618: contig of 1524 bp in length
* 23619 23718: gap of unknown length
* 23719 24845: contig of 1127 bp in length
* 24846 24945: gap of unknown length
* 24946 26453: contig of 1488 bp in length
* 26454 26533: gap of unknown length
* 26534 28392: contig of 1859 bp in length
* 28393 28492: gap of unknown length
* 28493 30382: contig of 1890 bp in length
* 30383 30482: gap of unknown length
* 30483 32223: contig of 1741 bp in length
* 32224 32323: gap of unknown length
* 32324 33841: contig of 1518 bp in length
* 33842 33941: gap of unknown length
* 33942 35473: contig of 1532 bp in length
* 35474 35573: gap of unknown length
* 35574 36870: contig of 1297 bp in length
* 36871 36970: gap of unknown length
* 36971 38600: contig of 1630 bp in length
* 38601 38700: gap of unknown length
* 38701 40028: contig of 1328 bp in length
* 40029 40128: gap of unknown length
* 40129 41493: contig of 1365 bp in length
* 41494 41593: gap of unknown length
* 41594 42911: contig of 1318 bp in length
* 42912 43011: gap of unknown length
* 43012 44283: contig of 1272 bp in length
* 44284 44383: gap of unknown length
* 44384 46493: contig of 2110 bp in length
* 46494 46593: gap of unknown length
* 46594 48001: contig of 1408 bp in length
* 48002 48101: gap of unknown length
* 48102 50018: contig of 1917 bp in length
* 50019 50118: gap of unknown length
* 50119 51837: contig of 1719 bp in length
* 51838 51937: gap of unknown length
* 51938 54876: contig of 2939 bp in length
* 54877 54976: gap of unknown length
* 54977 56610: contig of 1634 bp in length
* 56611 56710: gap of unknown length
* 56711 58795: contig of 2085 bp in length
* 58796 58895: gap of unknown length
* 58896 60962: contig of 2067 bp in length
* 60963 61062: gap of unknown length
* 61063 63317: contig of 2255 bp in length
* 63318 63417: gap of unknown length
* 63418 64835: contig of 1418 bp in length
* 64836 64935: gap of unknown length
```

```
* 64936 66869: contig of 1934 bp in length
* 66870 66969: gap of unknown length
* 66970 69172: contig of 2203 bp in length
* 69173 69272: gap of unknown length
* 69273 71451: contig of 2179 bp in length
* 71452 71551: gap of unknown length
* 71552 73345: contig of 1794 bp in length
* 73346 73445: gap of unknown length
* 73446 76108: contig of 2663 bp in length
* 76109 76208: gap of unknown length
* 76209 78870: contig of 2662 bp in length
* 78871 78970: gap of unknown length
* 78971 81769: contig of 2799 bp in length
* 81770 81869: gap of unknown length
* 81870 84942: contig of 3073 bp in length
* 84943 85042: gap of unknown length
* 85043 87086: contig of 2044 bp in length
* 87087 87186: gap of unknown length
* 87187 89756: contig of 2570 bp in length
* 89757 89856: gap of unknown length
* 89857 92029: contig of 2173 bp in length
* 92030 92129: gap of unknown length
* 92130 95030: contig of 2901 bp in length
* 95031 95130: gap of unknown length
* 95131 98584: contig of 3454 bp in length

Query Match      5.8%: Score 44; DB 2; Length 165337;
Best Local Similarity 42.5%: Pred. No. 48;
Matches 233; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

QY 62 TTGCTGCCGTGGCATTTTGCGCCGTTTGCGCTGATGCAATTCATGATATAAATACCGG 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104277 TTGTTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 104218

QY 122 AAGAAATCAACAAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104217 TTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 104158

QY 182 ATCCAAATGAAGATACCTGATGATGATGATGATGATGATGATGATGATGATGATG 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104157 CTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 104098

QY 242 TCAAGGTGATATTAGCCATGCGAAGCAATTCAGAGCTTGAGAGATTGAAACAAGACGTC 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104097 TTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 104038

QY 302 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104037 TTGTTGCTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 103978

QY 362 TTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103977 TTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 103918

QY 422 CAACCACTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103917 TTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 103858

QY 482 CTGATGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103857 TACTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 103798

QY 542 ACCATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103797 TTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 103738

QY 602 CCGCTATT 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103737 TTGCTGTT 103730

RESULT 26
AC091787
```


LOCUS	AC091187	170668 bp	DNA	linear	HTG-08-MAY-2002
DEFINITION	Oryza sativa chromosome 3 clone OSJNB0087611, *** SEQUENCING IN PROGRESS ***, 3 ordered pieces.				
ACCESSION	AC091187				
VERSION	AC091187.5	GI:20503084			
KEYWORDS	HTG; HTGS; PHASE2.				
SOURCE	Oryza sativa.				
ORGANISM	Oryza sativa.				
REFERENCE	Eukariota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzeae; Oryza. 1 (bases 1 to 170668)				
AUTHORS	Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Tallon,L., Gansberger,K., Kim,M., Overton II,L., Bera,D., Tslirih,T., Kroll,M., Jarrabhi,B., Jin,S., Koo,H., Zismann,V., Hsieh,J., Blunt,S., Vanaken,S., Uteckbach,T., Feildblum,T., Yang,Q., Haas,B., Suh,B., Peterson,J., Quackenbush,J., White,O., Salzberg,S. and Fraser,C. Oryza sativa ssp. japonica cv. Nipponbare OSJNB0087611 BAC genomic sequence				
TITLE	Unpublished 2 (bases 1 to 170668)				
JOURNAL	Buell,R.				
REFERENCE	Direct Submission Submitted (06-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
JOURNAL	Medical Center Dr, Rockville, MD 20850, USA 3 (bases 1 to 170668) Buell,R.				
REFERENCE	Direct Submission Submitted (08-MAY-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On May 8, 2002 this sequence version replaced gi.15987233.				
AUTHORS	* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes * of the gaps between them are based on estimates that have * provided by the submittor.				
TITLE	* This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.				
JOURNAL	1 101260: contig of 101260 bp in length * 101261 101360: gap of unknown length * 101361 120773: contig of 19413 bp in length * 120774 120873: gap of unknown length * 120874 170668: contig of 49795 bp in length.				
FEATURES	Location/Qualifiers 1..170668 /organism="Oryza sativa" /cultivar="Nipponbare" /db_xref="taxon:4530" /chromosome="3" /clone="OSJNB0087611" /note="japonica cultivar-group"				
SOURCE	BASE COUNT 49352 a 36238 c 36061 g 48553 t 464 others ORIGIN				
Query Match	5.8%; Score 44; DB 2; Length 170668; Best Local Similarity 20.6%; Pred. No. 48;				
Matches	68; Conservative 120; Mismatches 140; Indels 2; Gaps 1;				
Oy	3 TCTTATATCAATTAACAATCCAAAAAACAATATTCTTACAAATAGATAAATTTGGTGAT 62 Db 10710 TCMAACAACCTAAAGAAAATATCATCAGCAAAATTTWAAAACTATRKKKWSMAMWMMARW 10769 				
Oy	63 TGTCGCCGTGCAGATTGCGCGCGTTGCGCGATGCCAATTCGA -CTATGATTAATATCAC 120 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : Db 10770 TRAKAAMMYTRTWMTWRMRMAYYTKKRRAKMWAISYKMKWTMRSCWMARKMGAMAAT 10829 				
Oy	121 GAAGAATAACAACAAGCATTTGATGATGCCATTCGCTGATTTGAAACATCCGAACAATA 180 ::: : : : : : : : : : : : : : : : : : : : Db 10830 KMCRMYMTRRAYAWMRWRRRTKTGRKTYWWSMTWKMARBMKAGANTMRYAATCKKKK 10889 				
Oy	181 GATCCAATGAAAGTACTGATCATGCGCATTAATTCGAACGTCATGTTGGTATTTGGAT 240				

QY	DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	REFERENCE	TITLE	JOURNAL	COMMENT	FEATURES	source
QY 241	TTCAAAGGCTGAATTACGCATCGAACAATTGAGGCTGAGGATTTGAAACAATGATGAACACGTT	AC115578	8813 bp DNA linear HTG 21-MAR-2002	AC115578	1	GI:19569893	HTG: HTGS_PHASE2.	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum
QY 10950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950	110950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950														
QY 301	CAAGTGATGCTAATGTCTAAAGGTCGAGAG	330																	
Db 11010	WGAGKAWMSKMKSCWMMGAAAKTKYAYRS	11039																	
QY 27	AC115578/c	8813 bp	DNA linear HTG 21-MAR-2002	AC115578	1	GI:19569893	HTG: HTGS_PHASE2.	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum
QY 10950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950	110950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950														
QY 301	CAAGTGATGCTAATGTCTAAAGGTCGAGAG	330																	
Db 11010	WGAGKAWMSKMKSCWMMGAAAKTKYAYRS	11039																	
QY 27	AC115578/c	8813 bp	DNA linear HTG 21-MAR-2002	AC115578	1	GI:19569893	HTG: HTGS_PHASE2.	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum
QY 10950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950	110950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950														
QY 301	CAAGTGATGCTAATGTCTAAAGGTCGAGAG	330																	
Db 11010	WGAGKAWMSKMKSCWMMGAAAKTKYAYRS	11039																	
QY 27	AC115578/c	8813 bp	DNA linear HTG 21-MAR-2002	AC115578	1	GI:19569893	HTG: HTGS_PHASE2.	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum
QY 10950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950	110950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950														
QY 301	CAAGTGATGCTAATGTCTAAAGGTCGAGAG	330																	
Db 11010	WGAGKAWMSKMKSCWMMGAAAKTKYAYRS	11039																	
QY 27	AC115578/c	8813 bp	DNA linear HTG 21-MAR-2002	AC115578	1	GI:19569893	HTG: HTGS_PHASE2.	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum	Dictyostelium discoideum
QY 10950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950	110950	WMKGTGRCMAAARBMCMRYMMWRCGKAATAATWTYMMYMSKSKAAYCTGRTYSW	110950														
QY 301	CAAGTGATGCTAATGTCTAAAGGTCGAGAG	330																	
Db 11010	WGAGKAWMSKMKSCWMMGAAAKTKYAYRS	11039																	
QY 27	AC115578/c	8813 bp	DNA linear HTG 21-MAR-2002	AC115578	1	GI:19569893													

```

/codon_start=1
/product="HYPOPHYSICAL 18.6 KDA PROTEIN. 6/101"
/protein_id="AAL92211.1"
/db_xref="GI:19569895"
/translation="MSTDIYIGIIIGVILGYFOAKCFKSRITTKKEEVDIOKEIEN
NNQKEKEKEKEDDDDDDESDSESDIGDYDENDDYDEIYDEIDGCKYVIRFDLG
MKKGITACCCASIGAROKALKDSKRVLRWENFCGAXIXIKRANKSELDLRNKA
ELGVSYVLDSGHTQIPSGSATVLAVGAPKPEAVDAVAKHLKLY"
BASE COUNT      3563 a      920 c      934 g      3395 t      1 others
ORIGIN
Query Match      5.8%; Score 43.8; DB 2; Length 8813;
Best Local Similarity 53.9%; Pred. No. 81;
Matches 90; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY  99  AATTCATCATTAATACCGGAAATCAACAAAGCTATTGATGATCCATTCGTC 158
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  6144 AATTAATATCAAAAAGAAAAAGAAAAAGAAAAAGAGGATGATGATGATGATCA 6085
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  159  TATTGAACATCCGAAACATAGATCCAAATGAAGTACCTGATCAGCCGTAATTCGA 218
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  6084 TGATGATGAAGAAGTGAATGATGATGATGATGATGAAGAATGATGATGATGA 6025
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  219  ACCTCATGTTGTATTGTGATTTCAAGGTGAATTAACCATCGCA 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  6024 ATATGATGAAGAATGTGATTTGAACAAGTCTTGAATTAAGAA 5978
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 28
AC099368/c 166613 bp DNA linear HNG 12-JUL-2002
LOCUS      Rattus norvegicus clone CH230-84G11, *** SEQUENCING IN PROGRESS
DEFINITION *** 75 unordered pieces.
ACCESSION AC099368
VERSION   AC099368.3 GI:21728883
KEYWORDS  HNG; HNGS_PHASE1.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1 (bases 1 to 166613)
           Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
           Alsbrooks,S.L., Anaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
           Barbieri,J., Benton,J., Blimang,K., Blankenburg,K., Bonini,D.,
           Bouck,J., Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
           Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
           Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
           Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
           Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
           Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
           Douthwalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
           Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
           Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
           Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
           Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
           Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
           Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
           Homel,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
           Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
           Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
           Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
           Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,J., Loulseged,H.,
           Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
           Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
           Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
           Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
           Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
           Nguyen,N., Nickerson,E., Nwokwko,S., Oguh,M., Okunou,G.,
           Oraguine,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
           Peters,J., Pichens,R., Primus,E., Pu,J.L., Quiles,M., Ren,Y.,
           Rivers,M., Rojas,A., Rojudoxan,I., Rolfe,M., Ruiz,S., Savery,G.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 166613)
Worley,K.C.
Direct Submission
Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 166613)
Worley,K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:1793175.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc.help@bcm.tmc.edu
----- Project Information
Center project name: GHJ
Center clone name: CH230-84G11
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 90800 bases at least Q40
Consensus quality: 97795 bases at least Q30
Consensus quality: 102507 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1137: contig of 1137 bp in length
* 1138 1237: gap of unknown length
* 1238 2274: contig of 1037 bp in length
* 2275 2374: gap of unknown length
* 2375 4033: contig of 1659 bp in length
* 4034 4133: gap of unknown length
* 4134 5519: contig of 1386 bp in length
* 5520 5620 6939: gap of unknown length
* 6940 7040 8283: contig of 1244 bp in length
* 8284 8384: gap of unknown length
* 8385 9671: contig of 1288 bp in length
* 9672 9771: gap of unknown length
* 9772 11096: contig of 1335 bp in length
* 11097 11196: gap of unknown length
* 11197 12269: contig of 1073 bp in length
* 12270 12370 13670: gap of unknown length
* 13671 13770: contig of 1301 bp in length
* 13771 14791: gap of unknown length
* 14792 14891: contig of 1021 bp in length
* 14892 15917: gap of unknown length

```

```
* 15918 16017: gap of unknown length
* 17081 17084: contig of 1067 bp in length
* 17085 17184: gap of unknown length
* 18538 17185: contig of 1354 bp in length
* 18539 18638: gap of unknown length
* 20241 20241: contig of 1603 bp in length
* 20342 20341: gap of unknown length
* 20342 21950: contig of 1609 bp in length
* 21951 22050: gap of unknown length
* 22051 23699: contig of 1649 bp in length
* 23700 23799: gap of unknown length
* 23800 25374: contig of 1575 bp in length
* 25375 25474: gap of unknown length
* 25475 26817: contig of 1343 bp in length
* 26818 26917: gap of unknown length
* 26918 28206: contig of 1289 bp in length
* 28207 28306: gap of unknown length
* 28307 29757: contig of 1451 bp in length
* 29758 29857: gap of unknown length
* 29858 31524: contig of 1667 bp in length
* 31525 31624: gap of unknown length
* 31625 33275: contig of 1651 bp in length
* 33276 33375: gap of unknown length
* 33376 34845: contig of 1470 bp in length
* 34846 34945: gap of unknown length
* 34946 36552: contig of 1707 bp in length
* 36553 38274: gap of unknown length
* 38275 38374: gap of unknown length
* 38375 39946: contig of 1572 bp in length
* 39947 40046: gap of unknown length
* 40047 41323: contig of 1277 bp in length
* 41324 41423: gap of unknown length
* 41424 42560: contig of 1137 bp in length
* 42561 44328: gap of unknown length
* 44329 44428: contig of 1668 bp in length
* 44429 46434: gap of unknown length
* 46435 46534: gap of unknown length
* 46535 48108: contig of 1574 bp in length
* 48109 48208: gap of unknown length
* 48209 50094: contig of 1886 bp in length
* 50095 50194: gap of unknown length
* 50195 51220: contig of 1026 bp in length
* 51221 51320: gap of unknown length
* 51321 53218: contig of 1898 bp in length
* 53219 53318: gap of unknown length
* 53319 54394: contig of 1076 bp in length
* 54395 54494: gap of unknown length
* 54495 56577: contig of 2083 bp in length
* 56578 56677: gap of unknown length
* 56679 59069: contig of 2392 bp in length
* 59070 59169: gap of unknown length
* 59170 61368: contig of 2199 bp in length
* 61369 61468: gap of unknown length
* 61469 63423: contig of 1955 bp in length
* 63424 63523: gap of unknown length
* 63524 65368: contig of 1846 bp in length
* 65369 65469: gap of unknown length
* 65470 68413: contig of 2944 bp in length
* 68414 68513: gap of unknown length
* 70491 70491: contig of 1978 bp in length
* 70492 70591: gap of unknown length
* 71639 71639: contig of 1048 bp in length
* 71640 71739: gap of unknown length
* 71740 73434: contig of 1695 bp in length
* 73435 73534: gap of unknown length
* 73535 74959: contig of 1425 bp in length
* 74960 75059: gap of unknown length
* 75060 76903: contig of 1844 bp in length
* 76904 77003: gap of unknown length
* 77004 78655: contig of 1652 bp in length
* 78656 78755: gap of unknown length
```

```
* 78756 80321: contig of 1566 bp in length
* 80322 80421: gap of unknown length
* 80422 80953: contig of 2532 bp in length
* 82954 83053: gap of unknown length
* 83054 84194: contig of 1141 bp in length
* 84195 84294: gap of unknown length
* 84295 85312: contig of 1018 bp in length
* 85313 85412: gap of unknown length
* 85413 86858: contig of 1446 bp in length
* 86859 86958: gap of unknown length
* 86959 90750: contig of 3792 bp in length

Query Match      5.88; Score 43.8; DB 2; Length 166613;
Best Local Similarity 42.68; Pred. No. 53;
Matches 228; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

Qy 62 TTGCTGCCGTCGATTTGTCGCCGTTTCGCGTCGATTCCAATTCACATGATAAAATCACCG 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117188 TTGTTGTTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCG 117129

Qy 122 AAGAAATCAACAAAGCTATTCATGATGCCATTCGTCGTATTCGAACAAATCCGAACAAATAG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117128 ATGCTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 117069

Qy 182 ATCCATGAAGATACGTCATGATCCGATTAATTCGAACGTCATGTTGATTCGATTC 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117068 TTGCTATTTGTTGATGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 117009

Qy 242 TCAAAATGAATATAGCCATGCGCAATTCGAGCTCGAGATTCGAACAAATGAACATC 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117008 CTGTTATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 116949

Qy 302 AAGTGATGCTAATGCTCAAGGTCAGAGGTCATTTGAAGTCATTTGTCATTCGTCG 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116948 TTGCTGTTGTTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 116889

Qy 362 TTCAGATGATATGTCGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116888 TTGTTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 116829

Qy 422 CAACTGATGATTCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116828 CTGTTGCTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 116769

Qy 482 CTGATGAAGGTAACATTAACATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116768 TTGTTGTTGCTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 116709

Qy 542 ACCATATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116708 TTGCTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 116654

RESULT 29
AL162726/c 195824 bp DNA linear PRI 13-NOV-2001
LOCUS Human DNA sequence from clone RP11-15B24 on chromosome 9, complete
DEFINITION sequence.
ACCESSION AL162726
VERSION AL162726.16 GI:15722081
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Ashwell, R.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Sep 22, 2001 this sequence version replaced gi:15590843.
COMMENT During sequence assembly data is compared from overlapping clones.
```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EM: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human Chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>

Rp11-15B24 is from the library Rp11-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pACE3.6

This sequence is the entire insert of clone Rp11-15B24 The true left end of clone Rp11-33B6 is at 112569 in this sequence. The true right end of clone Rp11-432M2 is at 24329 in this sequence.

FEATURES

Source

1.195824
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="Rp11-15B24"
/clone.lib="Rp11-11.1"

BASE COUNT 60164 a 36668 c 38052 g 60940 t
ORIGIN

Query Match

Best Local Similarity 5.8%; Score 43.8; DB 9; Length 195824;

Matches 81; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 612 CCAGACACCGTACGTAAGGAATGACCAAGTATTGGCAGCAGATTTAAACGTGAATT 671

DB 153781 CAACCCACGACAGAAAAAATAAAGATCAAGCAGAGCAATGAATGAAC 153722

QY 672 GAAAAAATTAACATGACATCTTTTCCAACTGACATCTCTATTTCACGACA 731

DB 153721 AGAAAAAATATCAAAAGATTAATTAATTAAGTTGTATATTTCCCTTTGTTGCA 153662

QY 732 ATAAATTAATTTTATTTTAA 754

DB 153661 ATAAATTCAGCAATTAATTTTAA 153639

RESULT 30

AC096351/c 227145 bp DNA linear HTG 11-JUL-2002

LOCUS Rattus norvegicus clone CH230-24M6, *** SEQUENCING IN PROGRESS ***

DEFINITION 25 unordered pieces.

ACCESSION AC096351

VERSION AC096351.4 GI:21723519

KEYWORDS HTG; HTGS-PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 227145)

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alstrooms,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,U., Benton,J., Blumage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,L.P.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,X., Dim,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,T., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunarane,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Hayla,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korval,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,J., Liu,W., Louisegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rojfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshat,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tameisa,A., Tameisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Teitrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 227145)

Worley,K.C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 227145)

Worley,K.C.

Direct Submission

Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jul 10, 2002 this sequence version replaced gi:20976103.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GRZD

Center clone name: CH230-24M6

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 20552 bases at least Q40

Consensus quality: 20867 bases at least Q30

Consensus quality: 20963 bases at least Q20

NOTE:

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 25 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

Db 145192	TTTTGACACCCCTTTAGAACCTAGACAGGCGCTGGCATACACAGAAAGTTGTTGTTGCTT	145133
Oy 63	-TGTGCGCGTGGCAGATTTTGTGCGCGCTTTTCGGCTGATCCAAATTCAGTATGATATAAAATCAC	120
Db 145132	GCTGCTGCTGTTGTTGTTGCTGCTGCTGTTGTTGTTGCTGCTGCTGTTGTTGCTGCTGCTGCTGCT	145073
Oy 121	GAAAGAAATCAACAAAGCTATTGATGATGATGCCATTTGCTGCTATTGGAACAAATCCGAAACAATA	180
Db 145072	GTTGTTGTTGCTGTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	145013
Oy 181	GATCCCAATGAAGAATGATCATGATCATGCCGATATAATTCGAACGTCATGTTGATATGATGAT	240
Db 145012	GCTGCTGTTGTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	144953
Oy 241	TTCAAAAGGTGAATTAAGCCATGCGAAACATTGAGGCTCGAGAGATTGGAACAAATGAAACGT	300
Db 144952	GTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	144893
Oy 301	CAGGTGATGCTAATGCTCAAAAGGTGAAGAGGTATTGTTAAAGCTCATTTGTTGATCGCT	360
Db 144892	GCTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	144833
Oy 361	GTTTCACGATGATATGCTGCTCGATGGAATTAATGATTAAGATACAAATTTGGGATCTGAT	420
Db 144832	GCTGTTGTTGTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	144773
Oy 421	CCAACCACTCATGTCATTTTCGGAATTTTCAGATATTTTGTGTTGCCCTTTCCTCCGTAAT	480
Db 144772	GCTGCTGCTGTTGTTGTTGTTGCTGCTGCTGCTGCTGCTGTTGTTGTTGCTGCTGCTGCTGCTGCT	144713
Oy 481	TCTGATGAAGGTACATACATACATGACATCTTTGAAAGTACGACAAATTCGCTAATGTTGTC	540
Db 144712	GCTGCTGTTGTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	144653
Oy 541	AACCAATTTGGGATGCTTTCATCTTTGATATCCAAATTTTGGGCTGTTTTCATGATATATG	600
Db 144652	GTTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	144593
Oy 601	ACCGCTATTTT 611	
Db 144592	GTTGCTGTTGCT 144582	
RESULT 34		
PFMAL3P6	164399 bp DNA linear INV 05-MAY-2000	
LOCUS	PFMAL3P6	
DEFINITION	Plasmidium falciparum MAL3P6, complete sequence.	
ACCESSION	Z98551 AL010161 AL010170 AL010212 AL010213 AL022222 AL139179 Z98553	
VERSION	Z98551.1 GI:3738836	
KEYWORDS	HTG; 40S Ribosomal protein S11; 40S Ribosomal protein S15A; band 7-related protein; CDC2-related protein kinase; DNA-directed RNA polymerase II; FN transporter family; kinesin-related protein; P-type ATPase ring finger protein; PDZ domain; proteasome component C8; putative cleavage and polyadenylation specificity factor protein; putative inorganic pyrophosphatase; triosephosphate isomerase; trophozoite stage antigen; ubiquitin-conjugating enzyme; zinc-finger protein.	
SOURCE	Plasmidium falciparum 3D7.	
ORGANISM	Plasmidium falciparum 3D7.	
REFERENCE	Enuairi, A., Alveolates; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 164399)	
AUTHORS	Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Fellwell, T., Gentles, S., Gilliam, R., Hamlin, N., Harris, D., Holtroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Mould, S., Mungall, K., Murphy, U., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and Barrell, B.G.	
TITLE	The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum	
JOURNAL	Nature 400 (6744), 532-538 (1999)	
MEDLINE	99376085	

FEATURES	source
COMMENT	10448855 2 (bases 1 to 164399) Murphy, L., Bowman, S., Harris, D., Lawson, D., Quail, M. and Barrell, B. JOURNAL Unpublished 3 (bases 1 to 164399) Lawson, D., Bowman, S. and Barrell, B. TITLE Direct Submission Submitted (14-AUG-1997) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK On or before May 14, 2001 this sequence version replaced gi:2665320, gi:2982570, gi:2982576, gi:2894503, gi:2982580, gi:2894373, gi:2982530. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum . Location/Qualifiers 1..164399
gene	/organism="Plasmodium falciparum 3D7" /strain="3D7" /db_xref="taxon:36329" /chromosome="3" /clone="MAL3P6" complement(152..1075) /gene="PFC0700c, MAL3P6.1" complement(152..1075) /gene="PFC0700c, MAL3P6.1" /note="PFC0700c (MAL3P6.1), hypothetical protein, len: 307 aa, predicted using hexexon" /codon_start=1 /product="hypothetical protein, PFC0700c" /protein_id="CAB11159.1" /db_xref="GI:3738865" /db_xref="SP:SPREMBL:077394" /translation="MHDFELKSFNLISSPLFNNFYRNNEDEYFKKRRNNNDLGYM HNYADSEMRHNKKORMTSLKNEINOLITYYNNFNNNYENYKSTELKEKNND EYNEQOYEPTANLQDKNKINDNNNYNNFNKNSLFFNCONFOADKNPLYLANKKAK NNTNENLIDVEFKLNKHYFLDMHNVNNDQQLTWHLITPTKPMHHIINKPVS SSNAVSEIKNKKDKKNGTMYLLKTLKLSSEFNQMTFVSSIFISFLINKHWRALAKI SQLQKINSNFLKSVRLFEESSLIRKNKYI" complement(156..161) /gene="PFC0700c, MAL3P6.1" /note="potential splice donor sequence at 3' end of PFC0700c may indicate splicing" 2866..3762 /gene="PFC0701w" /note="PFC0701w, hypothetical protein, len: 298 aa, revised: new gene prediction" /codon_start=1 /product="hypothetical protein, PFC0701w" /protein_id="CAB90508.1" /db_xref="GI:7768284" /translation="MNTKLPFGNSLCKILKYNVNNLRGCIYNAAGPFIISYLKLEIT NNKKESSCSSNIGIKRGYGNKKRSFELKLRNFENYGNFCTMKIRMTNNINNY NSDSDNVNFKTFKRYKAVKTSYKRSKRSFSGNGQAKIGIKRLSGEYATGQMLAK ORAKICEYENKTRRRNFKYYPGENVYAKRTSLIATLNGNLPKTFVIONVFLVNLIL PEELIEKEDLDLYRTEHVKYSPEDNSLIYLRMKNNILFPKYKOTGYIRPLKPGFL TKYIYDNPITQKVPILYHKKE" complement(1080..7204,7332..11478) /gene="PFC0705c, MAL3P6.2" complement(1080..7204,7332..11478) /gene="PFC0705c, MAL3P6.2" /note="PFC0705c (MAL3P6.2), hypothetical protein, len: 2423 aa, predicted using hexexon but revised: changed splice donor sequence for exon 1" /codon_start=1 /product="hypothetical protein, PFC0705c" /protein_id="CAB11149.2" /db_xref="GI:7768285" /db_xref="SP:SPREMBL:077393" /translation="MTTFDIYNNKRLKKEKPCFKLRKSPSLSFQKCIEMCDIELIKRN IHDKRLAKLKYELKEIILYSILKTKRNFQLFHMEYNNKYNVETVTQYMFNLSM

Db 87988 AATAATGATACGAAAAATGATGACGTAGAAAAATGATGACGTAGAAAAATGATGA 88047

QY 183 TCCAAATGAAGTACCTGATCATGCGCATTAATTCGAACGTGATGGTATGTTGATTT 242

Db 88048 TGACGAAGAAAAATGATGATGATGAGAAAAATGATGACGACGAAAAATGATGATGA 88107

QY 243 CAAAGTGAATTTAGCCATGCGAAACATTGAGGCTCGAGATTGGAACAAATGAACGCTGA 302

Db 88108 AGAAAAATGATGATGACGTAGAAAAATGATGATGAGAAAAATGATGATGAGAAAA 88167

QY 303 AGGTGATGCTTAATGCTCAAGAGTGAAGCGTATTTGTTAAAGCTCATTTGTTGATCGGTG 362

Db 88168 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 88227

QY 363 TCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 422

Db 88228 TAAAGGAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 88287

QY 423 AACCACTCATGCTCATTTCCGATATTCAAGATTTTGTGTT 462

Db 88288 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 88327

RESULT 36

AC124960 128656 bp DNA linear HTG 23-AUG-2002

LOCUS medicago truncatula clone mth2-26j24, WORKING DRAFT SEQUENCE, 9

DEFINITION uncloned pieces.

ACCESSION AC124960

VERSION AC124960.6 GI:22450609

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE 1 (bases 1 to 128656)

AUTHORS Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

TITLE Cook,D., Kim,D. and Roe,B.A.

JOURNAL Medicago truncatula BAC Clone mth2-26j24

REFERENCE 2 (bases 1 to 128656)

AUTHORS Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

TITLE Cook,D., Kim,D. and Roe,B.A.

JOURNAL Direct Submission

Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT On Aug 23, 2002 this sequence version replaced gi:22038535.

----- Genome Center

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently

* consists of 9 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 2685: contig of 2685 bp in length

* 2686 2785: gap of unknown length

* 2786 6413: contig of 3628 bp in length

* 6414 6513: gap of unknown length

* 6514 13620: contig of 7107 bp in length

* 13621 13720: gap of unknown length

* 13721 23988: contig of 10268 bp in length

* 23989 24088: gap of unknown length

* 24089 34689: contig of 10601 bp in length

* 34690 34789: gap of unknown length

* 34790 46273: contig of 11484 bp in length

* 46274 46373: gap of unknown length

* 46374 60594: contig of 14221 bp in length

* 60595 60694: gap of unknown length

* 60695 89987: contig of 29293 bp in length

* 89988 90087: gap of unknown length

* 90088 128656: contig of 38569 bp in length.

FEATURES

source

1..128656

/organism="Medicago truncatula"

/db_xref="taxon:3880"

/clone="mth2-26j24"

/clone_1b="Medicago truncatula BAC library H2"

BASE COUNT 42345 a 19906 c 22362 g 43227 t 816 others

ORIGIN

Query Match

Best Local Similarity 48.4%; Pred. No. 71;

Matches 120; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 100 ATTTCATGATGATTAATACCCGAGAAATCAACAACCTATTGATGCCATTCGCT 159

Db 119167 ATGCCCTCAAGAACAAAGTTGATGGCATGATCATGATGAAGAAAGCAGCAGCA 119226

QY 160 ATTGAACAATCCGAACAATATGATCCATGAAGTACCTGATCCTGCCGATTAATTGAA 219

Db 119227 GAACAAAGACGCCGATGAGGAGGAGAAAGAAATCAACAGTTGATGGCATGATCATGAA 119286

QY 220 CGTCATGTTGATTTGATGATTTCAAGGTGAATTCACCATGATCCATGCAACATTTGACGCTGA 279

Db 119287 GATCAAGATGATTTATGATTAAGATGAAGATGATGAACACACACAGAAAGAGCGATGAA 119346

QY 280 GGATTTGAACAATGAAGCTCAAGGTGATGCTTAATGCTCAAGGCTGAAGAGCGTATTGTT 339

Db 119347 GATGAAGAAAGAAATGAACAAGTTGATGGCATGATCATGATGAAGATGATTAATGAT 119406

QY 340 AAGACTCA 347

Db 119407 AAGATGA 119414

RESULT 37

PCRAMP

LOCUS PCRAMP

DEFINITION P.chabaudi mRNA for a lysine-rich aspartic acid-rich protein, internal fragment.

ACCESSION X63619

VERSION X63619.1 GI:9805

KEYWORDS aspartic acid-rich protein; lysine-rich protein; repetitive sequence.

SOURCE Plasmodium chabaudi.

ORGANISM Plasmodium chabaudi.

REFERENCE 1 (bases 1 to 703)

AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

TITLE Hartz,D., Ayane,M., Chluba-De Tapie,J., Wirbelauer,C., Langhorne,J. and Gillard-Blass,S.

JOURNAL Cloning and sequencing of a cDNA fragment from Plasmodium chabaudi potentially that contains repetitive sequences coding for a parasitoid. Res. 79 (2), 133-139 (1993)

MEDLINE 93234423

PUBMED 8475031

REFERENCE 2 (bases 1 to 703)

AUTHORS Hartz,D.

JOURNAL Direct Submission

Submitted (23-DEC-1991) D. Hartz, Max-Planck-Institut fuer


```
/product="B9L"
/protein_id="AA69348.1"
/db_xref="GI:885733"
/translation="MFNYLENEVALDELKOMLKDRDPNDRNOFKNNALHAYLFENEH
CNNEEYAKLLDSGTNPLHKMRHPTPLEETNSRHYVKNKDIALMALLEAGFSNIN
FNINRYNKNVNDLKVLEHGFDSVCKENRSRYENVMTPDDPVEITIDFENN
GCNLYEDDEYGYDYKQPNCGTVLLYIISHLYSEDPTRAVRPPVCLINHG1
KPSIDKNYCTALQYIYKSSHIDIVLLMKGIDNTAYSYIDDLTRCTGIMADYLN
SDRYKNDVNDLKVLPLENGKPHGIMKSYPLRNRNKETISLILKTMNSVLDHILI
EYMFDDIDHILECMALYGC"
complement(7070..7366)
/codon_start=1
/codon="B10L"
/protein_id="AA69349.1"
/db_xref="GI:885734"
/translation="MQKEIADMCOIKKINGTDMLTVMKLNKPTKKRYVNNPIETEMTK
QOYKFYMOIYINANKLIEQSKRIDMINEVSDNNRSLTPELRHLIFSYYFL"
complement(7327..7845)
/codon_start=1
/codon="B11L"
/protein_id="AA69350.1"
/db_xref="GI:885735"
/translation="MPLEPVCVHNHDEDLVSLILCKPVIDIDINKIDTNGRSLLYYC
VYHNNTLLEWLVNGVDINATKCDYTCIGICIIILAYGCIPEIAKLYIOLECIL
SRLEPTIEKSTVENIKRKYVYIYKNSLIEMCIRYFILVNYKTYCDTSSYIEYTFE
CKKLLICVRLK"
7839..8120
/codon_start=1
/product="B12R"
/protein_id="AA69351.1"
/db_xref="GI:885736"
/translation="MAYNYCIIIEFYDINSIFONILNRYFYVYINKNHLRAVYIL
YNRPSSIIHILIVRTIPCEMYLVNNSMIEYLRLLIYVITTHIKOY"
complement(8406..8804)
/codon_start=1
/product="B13L"
/protein_id="AA69352.1"
/db_xref="GI:885737"
/translation="MLYDSCKTPDACSNOISVERENSNLYVTKKNNIKTIDVYVSL
LSTNYKINDFDIEYTESNIDVELRLIAGLEINSCKNGINIEKATTSNPN
VDVFKLLDQSIKPCMSNVSYGKIIIEKSY"
9039..9290
/codon_start=1
/product="B14R"
/protein_id="AA69353.1"
/db_xref="GI:885738"
/translation="MFVSCFNINGTNFLALCIIEVPTIKLYIOIMRCVYTCIDTA
HHLYIVYIOKFLKYLIREVYVSLGFKLSTVYGTSS"
complement(9454..9906)
/codon_start=1
/product="B15L"
/protein_id="AA69354.1"
/db_xref="GI:885739"
/translation="MGIOHEFDIIINGDIALRNLQLRGDNVCGCKLIISNDYKRLK
RFLIRPMSSEIDEVKGLTVFANNYAVKKNKDDPEFYIYEAVALHLYKKKTEILYSD
DENLFPKHVYVYISLNMISKYKVEEYSSPYIEHPLIRYOESMD"
complement(10136..10606)
/codon_start=1
/product="B16L"
/protein_id="AA69355.1"
/db_xref="GI:885740"
/translation="MNYNKRADSFSLSDSIKDYIHQYICWLSMTDEMRPSIGVFK
METFKIDAVRYDGNITDILADINAMSDSTIRSLONISSKDKKITVYGTNGLSIYV
DINKGCDISNKFAGIITILMEYIFPDNTMSHLVALYRIQRKRPIDDDDEDR"
complement(10962..11366)
/codon_start=1
/product="B17L"
/protein_id="AA69356.1"
/db_xref="GI:885741"
/translation="MDYVNRDLDDKLNHENSIGNITENAKKEVFGSRKITYDSVND
LIELNLAKYLNNEKLDVYERMDYVCKYIIGDMWTIYSFYKENGRLNSFLROYI
NNNIEIRNDFOPKFDVDSVDILINNDIETIV"
complement(11437..12387)
```

```
/codon_start=1
/product="B18L"
/protein_id="AA69357.1"
/db_xref="GI:885742"
/translation="MDYTIKIFNHGEFDTIRNKLVLNLLKVKWNTINSNVYVSTDTID
ISDCIRIELLYKOFKNVNIIEVSSNISFYKRNRPDPTLTDNVYVLYVLYNRKSVK
TLIYDIPREYVNTSSSEIDNEFSKSLNREPNRYDVKLVKSCSLSYKPSICIOYDNK
YIIDSDEGNLILCYVITMDPHLIDLETICVLVNSGKLSLNEFPYRBRKHNYD
SFADLCMDHIFELPDTBELPTLRDDGRNINAMDKLESSGNMTMIPRTDDEYKRLSKL
MNIAFNNTKFDYIMLVGDIDECTVFFFKVYKYYIINLME"
complement(12452..13243)
/codon_start=1
/codon="B19L"
/protein_id="AA69358.1"
/db_xref="GI:885743"
/translation="MKYERVTFLLTLLIGVLSCTTIPSRINNKFKNSVETDANAY
NIGDTIELCLPGYRKQKMPYIAKCTGTGTLNOCIKRRCSPRIDNCHLIDIGV
DFGSSITVSCNSGYLLIGYKSYCKLSTGSMVNNPKAPKADIESVCKOLPSISGRN
GYNDPYDGSVYVYSCNSGYSLIGNSGVLCGSGMSNPPTCOIKVCPPTILNGYLS
GFKRSYGVNDVDFCTKYGKLSGSSSSSTCSPGMTQPELPKVR"
complement(13311..13773)
/codon_start=1
/product="B20L"
/protein_id="AA69359.1"
/db_xref="GI:885744"
/translation="MGGVYVTFYKRYVILINDVCEPTKKNQWIKKSPMRHSIVAVED
GNITAITGITHETERNLYKRYIVKEDKNIELMYRNNHGMKMPVCCGDIYLIADAKTE
YPRKSNWNLNFMSTRNIEYIDMTFKDETPKCNVTHKSPSFLSNYEKOFLO"
```

Query Match 5.7%; Score 43; DB 14; Length 30288;
Best Local Similarity 46.3%; Pred. No. 96;
Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY	44	TGATGAATTTTGTGATGTCGCGTGCGATTTTGCGCGTTCGCTGATCCAATTC	103
DB	22856	TGCCGATGTTTATGCTAATATATCGTAGATATATAGATCATATGATATATAGATG	22797
QY	104	ACTATGATAAATCAACCGAAGAAATCAACAAAGCATATGATGATGCCAFTGCTGATTTG	163
DB	22796	ATATGATGATATATAGATGATATATAGATGATATATAGATATATAGATGATATAG	22737
QY	164	AACAATCCGAACATPAGATCCATGAAGTACCTGATCGATCGCATTAATTCGAACGTC	223
DB	22736	ATGATATATGATATATATAGATGATATATAGATATATAGATATATAGATATATAG	22677
QY	224	ATGTTGATTTGTCGATTTCAAGGTGAATTAGCCATCGGAACATTTAGCGTCGAGAT	283
DB	22676	TAGATGATATATGATATATATAGATGATATATAGATGATATATAGATGATATATAG	22617
QY	284	TGAACAAATGAACGTCAGGTGATGCTAATGTCAAGGTCGAAGGCGTATTGTTAAG	343
DB	22616	ATATGATGATAG	22557
QY	344	CTCATTT 350	
DB	22556	ATGATAT 22550	

RESULT 39
AC000117/c 110384 bp DNA linear PRI 21-DEC-1999
LOCUS
DEFINITION Homo sapiens BAC clone CTB-62A19 from 7q22, complete sequence.
ACCESSION AC000117
VERSION AC000117.1 GI:1809228
KEYWORDS
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 110384)
AUTHORS Magill, L. and Cordes, M.
TITLE The sequence of Homo sapiens BAC clone CTB-62A19
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 110384)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-1997)
REFERENCE 3 (bases 1 to 110384)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 110384)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Center: Washington University Genome Sequencing Center
Genome Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sepiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_RG062A19

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:
Clone CTB-62A19 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTA-317G18, 200 base overlap. Actual start of this clone is at base position 1 of CTB-62A19; actual end is at 110384 of CTB-62A19.

FEATURES
source This clone contains polymorphic bases with CTA-317G18.
1..110384 location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
/clone="CTB-62A19"
/clone_lib="CTB-978SK-B"
7..158
repeat_region
/rpt_family="Alu"
432..600
repeat_region
/rpt_family="L2"
487..634
SIS
/db_xref="GI:1113666"

repeat_region 2905..3218
/rpt_family="Alu"
repeat_region 3250..3374
/rpt_family="MIR"
repeat_region 3974..4058
/rpt_family="(TA)n"
repeat_region 4060..4370
/rpt_family="Alu"
4576..4641
repeat_region 5100..5495
/rpt_family="(TAA)n"
repeat_region 5941..6551
/rpt_family="Other"
repeat_region 6586..6724
/rpt_family="L2"
repeat_region 6790..7013
/rpt_family="MER1_type"
repeat_region 7028..7533
/rpt_family="MER1_type"
repeat_region 8537..8581
/rpt_family="L2"
repeat_region 8742..9037
/rpt_family="AT_rich"
repeat_region 9038..9185
/rpt_family="Alu"
repeat_region 9734..9879
/rpt_family="AT_rich"
repeat_region 9880..10061
/rpt_family="MER1_type"
repeat_region 10957..11706
/rpt_family="MIR"
repeat_region 11701..11796
/rpt_family="MER2_g"
repeat_region 12113..12781
/rpt_family="MER2_g"
repeat_region 12782..12846
/rpt_family="L1"
repeat_region 12847..13166
/rpt_family="Alu"
repeat_region 13179..13628
/rpt_family="L1"
repeat_region 13639..13784
/rpt_family="L1"
repeat_region 13806..14168
/rpt_family="L1"
repeat_region 14467..14897
/rpt_family="L1"
repeat_region 14900..14937
/rpt_family="L1"
repeat_region 14939..15173
/rpt_family="L1"
repeat_region 15174..15476
/rpt_family="L1"
repeat_region 15477..15900
/rpt_family="Alu"
repeat_region 18310..18607
/rpt_family="L1"
repeat_region 18886..19137
/rpt_family="Alu"
repeat_region 20169..20449
/rpt_family="MIR"
repeat_region 23034..23062
/rpt_family="Alu"
repeat_region 23940..24014
/rpt_family="AT_rich"
repeat_region 24015..24327
/rpt_family="L1"
repeat_region 25336..25648
/rpt_family="L2"
repeat_region 25906..26064
/rpt_family="Alu"
repeat_region 26757..26846
/rpt_family="L1"

```

repeat_region      /rpt_family="MER4-group"
26879..27213
repeat_region      /rpt_family="MER4-group"
27214..27643
repeat_region      /rpt_family="MALR"
27644..27695
repeat_region      /rpt_family="MER4-group"
28149..28339
repeat_region      /rpt_family="MIR"
29042..29339
repeat_region      /rpt_family="Alu"
30045..30105
repeat_region      /rpt_family="MER2-type"
30098..30596
repeat_region      /rpt_family="MER2-type"
30731..32294
repeat_region      /rpt_family="L1"
33461..34596
repeat_region      /rpt_family="L1"
34597..34960
repeat_region      /rpt_family="MALR"
34961..37079
repeat_region      /rpt_family="L1"
37080..37454
repeat_region      /rpt_family="MALR"
37457..39043
repeat_region      /rpt_family="MALR"
39044..39410
repeat_region      /rpt_family="MALR"
39411..39752
repeat_region      /rpt_family="L1"
40978..41021
repeat_region      /rpt_family="MIR"
42862..442561
repeat_region      /rpt_family="Alu"
44138..44352
repeat_region      /rpt_family="MIR"
45003..45190
repeat_region      /rpt_family="MIR"
45217..45301
repeat_region      /rpt_family="MIR"

Query Match      5.7%; Score 43; DB 9; Length 110384;
Best Local Similarity 50.2%; Pred. No. 79;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 390 TGATTACATACAAATGGGTGATCTTCATCCACACATCATGTCATTGGATATTCA 449
Db 12639 TGTTAAGCTTTGATTAATTTGTTTCAAAAAATTTGTCATCTCAAGTTGTTGA 12580
QY 450 AGATTTGTGTGGCTTGCTCCCTGGAATTTCTGATGAGGTACATTAACATGACATC 509
Db 12579 ATTTATTTACACAGACTGTTGCTATCTATTTATTTATTTATTTATTTATCTGAAATC 12520
QY 510 TTTTGAAGTACAGACAAATTCGCTAATGTTGTCACACCATATTGGTCTTTCAATTCGA 569
Db 12519 TGTAGTATATATTTCTTTCTTTGATGATGCTGATGATATATATTTCTGTTTCTTGCA 12460
QY 570 TCCAAATTTTGGCGTTTATCTGATGATG 600
Db 12459 CAGTTTGTAGTAGATTTAAACAATTTGTTG 12429

```

```

RESULT 40
AC012147      AC012147      217034 bp      DNA      linear      ROD 07-APR-2000
DEFINITION    Mus musculus, clone RP23-191A4, complete sequence.
ACCESSION     AC012147
VERSION       AC012147.7  GI:7523819
KEYWORDS      HMG.
SOURCE        Mus musculus.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS       Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE         Unpublished
JOURNAL       2 (bases 1 to 217034)
REFERENCE     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
AUTHORS       Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
               Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
               Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
               Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
               Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
               Lehoczy,J., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
               McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
               Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
               Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
               Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE         Direct Submission
JOURNAL       Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
AUTHORS       Research, 320 Charles Street, Cambridge, MA 02141, USA
               3 (bases 1 to 217034)
REFERENCE     Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
AUTHORS       Boguslavsky,L., Boukhalter,B., Brown,A., Burett,G.,
               Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
               Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,
               Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
               Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
               Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
               Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
               Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J.,
               Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
               McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
               Meldrum,J., Menes,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
               Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
               O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
               Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
               Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
               Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
               Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE         Direct Submission
JOURNAL       Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
AUTHORS       Research, 320 Charles Street, Cambridge, MA 02141, USA
               On Apr 7, 2000 this sequence version replaced gi:7342336.
COMMENT       All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ----- Genome Center
               - Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: MIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Center Project name: L906
               Center Clone name: 191_A_4
               -----
               FEATURES
               source
               location/Qualifiers
               1..217034
               /organism="Mus musculus"
               /db_xref="taxon:10090"
               /clone="RP23-191A4"
               /clone_1ib="RPCI-23 Female Mouse BAC"
               repeat_region
               2366..2394
               /rpt_family="(TTTGG)n"
               repeat_region
               2658..2999
               /rpt_family="L2"
               repeat_region
               3695..3793
               /rpt_family="(CAA)n"

```

```
repeat_region      3810..3906      /rpt_family="MER20"
repeat_region      4962..5094      /rpt_family="CT-rich"
repeat_region      complement(5097..5214)
repeat_region      /rpt_family="B1-F"
repeat_region      6052..6235      /rpt_family="B3"
repeat_region      6674..6741      /rpt_family="B3"
repeat_region      6751..6810      /rpt_family="(TA)n"
repeat_region      complement(7639..9521)
repeat_region      /rpt_family="Lx2"
repeat_region      complement(9650..16049)
repeat_region      /rpt_family="L1_MM"
repeat_region      complement(15990..16077)
repeat_region      /rpt_family="L1F"
repeat_region      complement(19061..19276)
repeat_region      /rpt_family="B3"
repeat_region      19303..19329      /rpt_family="AT-rich"
repeat_region      23171..23202      /rpt_family="(TC)n"
repeat_region      complement(23409..23523)
repeat_region      /rpt_family="ORR1B"
repeat_region      23524..23913      /rpt_family="MRA"
repeat_region      complement(23914..24015)
repeat_region      /rpt_family="ORR1B"
repeat_region      24016..24068      /rpt_family="GA-rich"
repeat_region      complement(24069..24265)
repeat_region      /rpt_family="ORR1D"
repeat_region      24638..24665      /rpt_family="(CAAA)n"
repeat_region      complement(25358..25583)
repeat_region      /rpt_family="Lx4"
repeat_region      25586..25883      /rpt_family="Lx4"
repeat_region      27522..27685      /rpt_family="Lx4"
repeat_region      /rpt_family="(TCTCC)n"
repeat_region      30456..30600      /rpt_family="MIR"
repeat_region      30746..30813      /rpt_family="AT-rich"
repeat_region      complement(31650..31718)
repeat_region      /rpt_family="ID3"
repeat_region      32388..32450      /rpt_family="(TTTT)n"
repeat_region      complement(32431..32565)
repeat_region      /rpt_family="B1_MM"
repeat_region      complement(32589..32660)
repeat_region      /rpt_family="MER63"
repeat_region      33234..33438      /rpt_family="Lx2"
repeat_region      33987..34081      /rpt_family="Lx9"
repeat_region      34086..34137      /rpt_family="(TAG)n"
repeat_region      34163..34255      /rpt_family="Lx9"
repeat_region      complement(34367..34649)
repeat_region      /rpt_family="ORR1A3"
repeat_region      34652..34708      /rpt_family="(TG)n"
repeat_region      complement(35105..35310)
repeat_region      /rpt_family="B3A"
repeat_region      37006..37038      /rpt_family="AT-rich"
repeat_region      complement(37793..38476)
repeat_region      /rpt_family="Lx9"
repeat_region      complement(38490..38934)
```

```
repeat_region      /rpt_family="L1"
repeat_region      38934..40250      /rpt_family="Lx"
repeat_region      40434..40484      /rpt_family="(CA)n"
repeat_region      40485..40603      /rpt_family="L1M1"
repeat_region      40749..40806      /rpt_family="(TC)n"
repeat_region      41007..41065      /rpt_family="(TG)n"
repeat_region      43582..43616      /rpt_family="(GA)n"
repeat_region      complement(43617..43842)
repeat_region      /rpt_family="URR1B"
repeat_region      44004..44017      /rpt_family="B4"
repeat_region      44018..44155      /rpt_family="B1_MM"
repeat_region      44160..44254      /rpt_family="(CAAA)n"
repeat_region      44326..44378      /rpt_family="(TAGA)n"
repeat_region      44372..44432      /rpt_family="(CAGA)n"
repeat_region      complement(44745..44912)
repeat_region      /rpt_family="L1"
repeat_region      complement(44949..45529)
repeat_region      /rpt_family="L1_MM"
repeat_region      45524..46469      /rpt_family="L1"
repeat_region      46521..46581      /rpt_family="AT-rich"
repeat_region      complement(47173..48057)
repeat_region      /rpt_family="L1M1"
repeat_region      48126..48281      /rpt_family="URR1B"
repeat_region      48283..48547      /rpt_family="Lx7"
repeat_region      48550..48679      /rpt_family="(TCTCC)n"
repeat_region      complement(48626..49035)
repeat_region      /note="Single-stranded terminator coverage."
repeat_region      48680..48857      /rpt_family="(TCTCC)n"
repeat_region      49041..49654

Query Match      5.7%; Score 43; DB 10; Length 217034;
Best Local Similarity 54.8%; Pred. No. 72;
Matches 85; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY      394 TTGACATCAAAATTTGGGTGATCTTCATCCAAACACATGTCATTTCGATATTCAGAT 453
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186910 TTTCCTTTCTTTTGGAGTGCTGACCTTCATGCCAATATGTAGTTTTTTGTAGACATT 186969
QY      454 TTTGTTGTGCTTGTCCCTTGAATTTTGATGAAGTAACATTAACATGACATCTTTT 513
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186970 TCTGTTTCCTCTTAATATGATTAATGATTAACAGATACATGTAATCATATCTTGA 187029
QY      514 GAAGTACGACAATTCGCTAATGTTGCAACCATAT 548
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187030 TATGACCACTTGAAAAATGTTGATATATCATAT 187064

RESULT 41
AC015797      AC015797      219294 bp      DNA      linear      HTG 09-MAY-2001
LOCUS      Mus musculus clone RP23-191A4, WORKING DRAFT SEQUENCE, 13 unordered
DEFINITION      pieces.
ACCESSION      AC015797.3      GI:8096845
VERSION      HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS      Mus musculus.
```



```

ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE          1 (bases 1 to 231739)
JOURNAL        Sloan,D., Hua,A. and Roe,B.A.
REFERENCE      Homo sapiens BNC Clone B250a12
AUTHORS        Unpublished
TITLE          2 (bases 1 to 231739)
JOURNAL        Sloan,D., Hua,A. and Roe,B.A.
REFERENCE      Direct Submission
TITLE          Submitted (15-OCT-1999) Department of Chemistry And Biochemistry,
JOURNAL        The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                OK 73019, USA
REFERENCE      3 (bases 1 to 231739)
AUTHORS        Sloan,D., Hua,A. and Roe,B.A.
TITLE          Direct Submission
JOURNAL        Submitted (12-OCT-2000) Department of Chemistry And Biochemistry,
                The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                OK 73019, USA

COMMENT        On Oct 11, 2000 this sequence version replaced gi:10281412.
FEATURES
    source
        1. .231739
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="unknown"
            /clone="B250a12"

BASE COUNT    56706 a 58574 c 58390 g 58069 t
ORIGIN
Query Match   5.7%: Score 43; DB 9; Length 231739;
Best Local Similarity 44.9%: Pred. No. 71;
Matches 163; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 134 AACGATTTGATGATGCCATTGCTCTATTGTGAACAATCCGAACAATAGATCAATGAAG 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184449 ATGATGTGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 184308
QY 194 TACCTGATGATGCCGATTAATTCGAACGTCATTTGGTATTGTTGATTTCAAGAGT 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1844309 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 184368
QY 254 TAGCCATGCCAATGATGAGGCTGAGAGTTGAACAATGAACGTCAGAGTATGCTA 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184369 ATGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 184428
QY 314 ATGCAAGAGTGAAGAGGATTTGTTAAAGCTCAATTTGTTGATCGGCTTCACAGTATA 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184429 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 184488
QY 374 TCCTCTGATGATGATGATTTAGCATACAAATTTGGGTATCTTCATCCACATCATG 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184489 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 184548
QY 434 TCATTTGGATTCACAGATTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184549 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 184608
QY 494 ACA 496
    |||
Db 184609 ATA 184611

RESULT 43
AC115288
LOCUS         AC115288 259940 bp DNA linear HTG 28-MAY-2002
DEFINITION   Mus musculus chromosome UNK clone RP23-47P24, WORKING DRAFT
SEQUENCE     AC115288
VERSION      AC115288.2 GI:21218537
KEYWORDS     HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE       Mus musculus.
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS        1 (bases 1 to 259940)
TITLE          McPherson,J.D. and Waterston,R.H.
JOURNAL        The sequence of Mus musculus clone
REFERENCE      Unpublished
AUTHORS        2 (bases 1 to 259940)
TITLE          McPherson,J.D. and Waterston,R.H.
JOURNAL        Direct Submission
REFERENCE      Submitted (16-MAR-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS        Parkway, St. Louis, MO 63108, USA
TITLE          3 (bases 1 to 259940)
JOURNAL        McPherson,J.D. and Waterston,R.H.
REFERENCE      Direct Submission
AUTHORS        Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
TITLE          Parkway, St. Louis, MO 63108, USA
JOURNAL        On May 28, 2002 this sequence version replaced gi:19526160.

COMMENT        ----- Genome Center -----
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc/index.shtml
                Contact: submissions@wustl.wustl.edu
                ----- Project Information -----
                Center project name: M_BA0047P24
                ----- Summary Statistics -----
                Sequencing vector: M13; 0%
                Sequencing method: plasmid; 100%
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 254498 bases at least Q40
                Consensus quality: 256163 bases at least Q30
                Insert size: 167000; agarose-fp
                Insert size: 258840; sum-of-ctnigs
                Quality coverage: 12.45 in Q20 bases; agarose-fp
                Quality coverage: 10.57 in Q20 bases; sum-of-ctnigs
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 12 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                1 16129: contig of 16129 bp in length
                * 16130 16229: gap of unknown length
                * 45224 45224: contig of 28995 bp in length
                * 45325 45324: gap of unknown length
                * 74631 74631: contig of 29307 bp in length
                * 74732 74731: gap of unknown length
                * 108761 108760: contig of 34029 bp in length
                * 108860 108860: gap of unknown length
                * 143821 143821: contig of 34961 bp in length
                * 143922 143921: gap of unknown length
                * 143922 182775: contig of 38854 bp in length
                * 182776 182775: gap of unknown length
                * 182876 182875: gap of unknown length
                * 234200 234199: contig of 51324 bp in length
                * 234200 234299: gap of unknown length
                * 235394 235394: contig of 1095 bp in length
                * 235395 235394: gap of unknown length
                * 235495 237306: contig of 1812 bp in length
                * 237307 237306: gap of unknown length
                * 237407 240876: contig of 3470 bp in length
                * 240877 240976: gap of unknown length
                * 240977 248528: contig of 7452 bp in length
                * 248529 248528: gap of unknown length
                * 248529 259940: contig of 11412 bp in length.
                -----
                Location/Qualifiers
                1. .259940
                /organism="Mus musculus"
FEATURES
    source

```

misc_feature /db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-47P24"
1..16129
misc_feature /note="assembly_name:Contig10"
16330..45224
/note="assembly_name:Contig11"
45325..74631
/note="assembly_name:Contig12"
74732..108760
/note="assembly_name:Contig13"
108861..143821
/note="assembly_name:Contig14"
143922..182775
/note="assembly_name:Contig15"
182876..234199
misc_feature /note="assembly_name:Contig16"
234300..235394
/note="assembly_name:Contig4"
235495..237306
/note="assembly_name:Contig5"
237407..240876
/note="assembly_name:Contig7"
240977..248428
misc_feature /note="assembly_name:Contig8"
248529..259940
/note="assembly_name:Contig9"
BASE COUNT 84077 a 48286 c 47975 g 78430 t 1172 others
ORIGIN

Query Match 5.7%: Score 43; DB 2; Length 259940;
Best Local Similarity 52.5%: Pred. No. 70;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Y 578 TTGGCGTTTATCTGATCTATTGACCGCTATTTCACAGACCGCTACGTAGAAGATGA 637

Db 172491 TTGGAAATGTAAGTAAATTAATACCTAATTAATAAAGAAATGTAATGAAGAAATAT 172550

Y 638 CCAAGATTTGGCACCACCATTTTAACGTGAATTGGAAGAAAAATTAACCAATAGCATCA 697

Db 172351 CTAAGAAATTAAGTAAATTTTAATAGAAATTAATAAATGAAGAAAAATTAAT 172610

Y 698 TTTTCCAACTGACATCTATTTTCACGTCAATTAATTAATTTTATTTTAT 756

Db 172611 ACTGTTATATGATACATGCTATATTAATTAATAACTTAGAAATGTGCATTTTAT 172669

RESULT 44

AC116031 14261 bp DNA linear HTG 23-MAR-2002

LOCUS Dictyostelium discoideum chromosome 2 map 140393-154652 strain AX4,
DEFINITION *** SEQUENCING IN PROGRESS ***, in ordered pieces.

ACCESSION AC116031.1 GI:19697368

VERSION HTG: HTGS_PHASE2.

KEYWORDS Dictyostelium discoideum.

SOURCE Dictyostelium discoideum

ORGANISM Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.

REFERENCE 1 (bases 1 to 14261)

AUTHORS Gloeckner,G., Etchlinger,T., Szafianski,K., Pachebat,J., Dear,P.,
Lehmann,R., Baumgart,C., Parra,G., Aprill,J.F., Guigo,R., Kumpf,K.,
Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
Noegel,A.A.

TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium

JOURNAL Unpublished

REMARK The Dictyostelium Genome Sequencing Consortium

REFERENCE 2 (bases 1 to 14261)

AUTHORS Baumgart,C.

JOURNAL Direct Submission

COMMENT Submitted (23-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
CDS predictions from Genaid may contain errors. Further information
is available from IMB Jena, Department of Genome Analysis

(http://genome.imb.jena.de/dictyostelium/
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS


```

repeat_region /rpt_family="L1"
541..714
/rpt_family="L1"
repeat_region /rpt_family="L1"
715..1012
/rpt_family="Alu"
1013..1180
/rpt_family="L1"
repeat_region /rpt_family="MIR"
1225..1394
/rpt_family="MIR"
1749..1828
/rpt_family="MALR"
1997..2219
/rpt_family="MALR"
2220..2519
/rpt_family="Alu"
2520..2573
/rpt_family="MALR"
2607..2953
/rpt_family="ERV"
2954..3301
/rpt_family="MALR"
3302..3377
/rpt_family="ERV"
3700..3988
/rpt_family="MALR"
3989..4280
/rpt_family="Alu"
4281..4363
/rpt_family="MALR"
4839..5017
/rpt_family="MIR"
5019..5042
/rpt_family="(TGGG)n"
5111..5128
/rpt_family="L1"
5129..5280
/rpt_family="L1"
5281..5703
/rpt_family="MALR"
5704..5856
/rpt_family="L1"
5857..6155
/rpt_family="Alu"
6156..6283
/rpt_family="L1"
6282..6505
/rpt_family="MER1_type"
6880..7265
/rpt_family="MALR"
8245..8387
/rpt_family="CT-rich"
8391..8743
/rpt_family="MALR"
8825..9232
/rpt_family="MALR"
9906..10207
/rpt_family="Alu"
10218..10382
/rpt_family="CT-rich"
10387..10529
/rpt_family="(TTCC)n"
10555..10806
/rpt_family="Alu"
11017..11375
/rpt_family="MALR"
11407..11612
/rpt_family="ERV"
11613..11926
/rpt_family="Alu"
11927..12186
/rpt_family="ERV"
12300..12473
/rpt_family="MIR"

```

```

repeat_region 12737..13035
/rpt_family="Alu"
repeat_region 13107..13283
/rpt_family="MIR"
repeat_region 13513..13770
/rpt_family="L2"
repeat_region 14040..14175
/rpt_family="ERV"
repeat_region 14903..15215
/rpt_family="Alu"
repeat_region 15329..15358

```

Query Match 5.6%; Score 42.8; DB 9; Length 16667;
 Best Local Similarity 43.4%; Pred. No. 82;
 Matches 197; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

```

QY 43 ATGATGAATTTTGTGATGCTGCGCGCATTTGTGCGCTTGGCGTATCCAAAT 102
Db 75525 ATGATGATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 75584
QY 103 CACTATGATTAATCAACCGAATAATCAACAAAGCTATTGATGATGCCATTGCTATT 162
Db 75585 AATGCGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75644
QY 163 GAACAAATCCGAACAAATAGATCCAAATGAAGTACATGATCCGATTAATTTGAACGT 222
Db 75645 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75704
QY 223 CATGTTGATTTGTGATTTCAAGTGAATTTGCCATGCAAAACATTTGAGCTCAGGA 282
Db 75705 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75764
QY 283 TTGAACAAATGAACGTCAGTGAATGCTAATGTCAAAGTGAAGGATTTGTTAA 342
Db 75765 GCGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75824
QY 343 GCTCATTTTGTGATCGGTTCACGATGATATGCTCGATGGAATATTTAGCATAC 402
Db 75825 GATGATGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75884
QY 403 AATTTGGTATCTTCATCCAAACCTCATGTCATTTGGATTTTGAATTTTGTGTT 462
Db 75885 GGTGATGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75944
QY 463 GCCTTGTCCCTGGAATTTCTGATGAAGTAA 496
Db 75945 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75978

```

```

RESULT 47
AC073449 75950 bp DNA linear HTG 18-JUN-2000
AC073449 Homo sapiens clone RP11-154A7, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC073449
ACCESSION AC073449
VERSION AC073449.1 GI:8571740
KEYWORDS HTG; HTGS; PHASE0.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 75950)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-154A7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 75950)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bede,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Cullymore,A., Cooke,P., Deatrelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,W., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,

```

TITLE
JOURNAL

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazars, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McSheeters, R., Meltrin, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pissani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tlirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zalnoun, J., Zimmer, A. and Zody, M.

Submitted (18-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L10390

Center clone name: 154_A_7

* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 734 833: gap of 100 bp
* 834 1540: contig of 707 bp in length
* 1541 1640: gap of 100 bp
* 1641 2334: contig of 694 bp in length
* 2335 2434: gap of 100 bp
* 2435 3156: contig of 722 bp in length
* 3157 3256: gap of 100 bp
* 3257 3981: contig of 725 bp in length
* 3982 4081: gap of 100 bp
* 4082 4828: contig of 747 bp in length
* 4829 4928: gap of 100 bp
* 4929 5658: contig of 730 bp in length
* 5659 5758: gap of 100 bp
* 5759 6493: contig of 735 bp in length
* 6494 6593: gap of 100 bp
* 6594 7329: contig of 736 bp in length
* 7330 7429: gap of 100 bp
* 7430 8169: contig of 740 bp in length
* 8170 8269: gap of 100 bp
* 8270 8974: contig of 705 bp in length
* 8975 9074: gap of 100 bp
* 9075 9806: contig of 732 bp in length
* 9807 9906: gap of 100 bp
* 9907 10636: contig of 730 bp in length
* 10637 10736: gap of 100 bp
* 10737 11426: contig of 690 bp in length
* 11427 11526: gap of 100 bp
* 11527 12261: contig of 735 bp in length
* 12262 12361: gap of 100 bp
* 12362 13097: contig of 736 bp in length
* 13098 13197: gap of 100 bp
* 13198 13938: contig of 741 bp in length
* 13939 14038: gap of 100 bp

14039 14779: contig of 741 bp in length
14780 14879: gap of 100 bp
14880 15624: contig of 745 bp in length
15625 15724: gap of 100 bp
15725 16471: contig of 747 bp in length
16472 16571: gap of 100 bp
16572 17304: contig of 733 bp in length
17305 17404: gap of 100 bp
17405 18113: contig of 709 bp in length
18114 18213: gap of 100 bp
18214 18940: contig of 727 bp in length
18941 19040: gap of 100 bp
19041 19753: contig of 713 bp in length
19754 19853: gap of 100 bp
19854 20551: contig of 698 bp in length
20552 20651: gap of 100 bp
20652 21378: contig of 727 bp in length
21379 21478: gap of 100 bp
21479 22207: contig of 729 bp in length
22208 22307: gap of 100 bp
22308 23029: contig of 722 bp in length
23030 23129: gap of 100 bp
23130 23872: contig of 743 bp in length
23873 23972: gap of 100 bp
23973 24699: contig of 727 bp in length
24700 24799: gap of 100 bp
24800 25534: contig of 735 bp in length
25535 25634: gap of 100 bp
25635 26371: contig of 737 bp in length
26372 26471: gap of 100 bp
26472 27211: contig of 740 bp in length
27212 27311: gap of 100 bp
27312 28018: contig of 707 bp in length
28019 28118: gap of 100 bp
28119 28842: contig of 724 bp in length
28843 28942: gap of 100 bp
28943 29675: contig of 733 bp in length
29676 29775: gap of 100 bp
29776 30499: contig of 724 bp in length
30500 30599: gap of 100 bp
30600 31314: contig of 715 bp in length
31315 31414: gap of 100 bp
31415 32126: contig of 712 bp in length
32127 32226: gap of 100 bp
32227 32948: contig of 722 bp in length
32949 33048: gap of 100 bp
33049 33774: contig of 726 bp in length
33775 33874: gap of 100 bp
33875 34596: contig of 722 bp in length
34597 34696: gap of 100 bp
34697 35415: contig of 719 bp in length
35416 35515: gap of 100 bp
35516 36255: contig of 740 bp in length
36256 36355: gap of 100 bp
36356 37068: contig of 713 bp in length
37069 37168: gap of 100 bp
37169 37900: contig of 732 bp in length
37901 38000: gap of 100 bp
38001 38732: contig of 732 bp in length
38733 38832: gap of 100 bp
38833 39559: contig of 727 bp in length
39560 39659: gap of 100 bp
39660 40389: contig of 730 bp in length
40390 40489: gap of 100 bp
40490 41219: contig of 730 bp in length
41220 41319: gap of 100 bp
41320 42061: contig of 742 bp in length
42062 42161: gap of 100 bp
42162 42900: contig of 739 bp in length
42901 43000: gap of 100 bp
43001 43747: contig of 747 bp in length
43748 43847: gap of 100 bp
43848 44593: contig of 746 bp in length

COMMENT CDS predictions from GeneID may contain errors. Further information

is available from IMB Jena, Department of Genome Analysis

Location/Qualifiers
1. .268147

BASE COUNT	99082 a	4117 c	76119 g	170662 t
ORIGIN				

Query Match	5.6%	Score 42.6	DB 6	length 349980
Best Local Similarity	54.9%	Pred. No. 80		
Matches 84	Conservative 0	Mismatches 69	Indels 0	Gaps 0

Dy 606 TATTTCGAGACACCCTAGCTAAGCAATGACCAAGTATTGGCACCCAGCATTTAAAC 665
||| ||| ||| ||| ||| ||| ||| |||
Db 291738 TAATAACCATTAACCTCTAAAAAAAATTAATAATCATTCTTCTAAAAAATCTTAAATC 291679

Dy 666 TGAATGGAAAAAATTAAACCATAGACATCAATTTTCCAACTGTCACAATCTCATTTTCA 725
+ + + + +
Db 291678 TAAATTTATTAACCTAAATAAATCTCACATTTTCTACTCTGTATTTTAACCACA 291619

Qy 726 CTGACATTAATAATTTTATTTTATTTTC 758
| | | | | | | | | | | |
Db 291618 ATTACTTAACACAAACCTTAAAAATCTTTTC 291586

Search completed: February 20, 2003, 21:52:34
Job time : 5701 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 19:25:16 : Search time 240 Seconds
(without alignments)
7140.711 Million cell updates/sec

Title: US-10-024-955-6

Perfect score: 761
Sequence: 1 GATCTTATATCAATACAAAT.....ATTTCATTTTATTTTCGCC 761

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.GeneSeq_101002:*

- 1: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	761	100.0	761	15	AAO71401
2	517	67.9	812	15	AAO71400
3	47.8	6.3	1386	24	ABN93011
4	42	5.5	7309	22	AA546567
5	42	5.5	7309	24	ABK34010
6	42	5.5	7309	24	ABL33816
7	41.8	5.5	16439	24	ABL32887
8	41.6	5.5	1367	23	ABL19335
9	41.6	5.5	4958	23	ABL19334

10	41	5.4	484	24	AA521133
11	40.6	5.3	11871	20	AA13108
12	40.2	5.3	4677	21	AA470259
13	40	5.3	14429	21	AB067097
14	40	5.3	14429	24	ABL34242
15	40	5.3	2365589	24	ABA90521
16	39.6	5.2	1702	19	AAV52500
17	39.6	5.2	2440	24	AA595258
18	39.6	5.2	32768	19	AAV52204
19	39.4	5.2	276	22	ABN91066
20	39.2	5.2	774	22	AAH52366
21	39.2	5.2	1194	22	AAH52925
22	39.2	5.2	3238	22	AAH54786
23	39.2	5.2	3931	22	AAH54467
24	39.2	5.2	4241	22	AAH34213
25	39	5.1	1350	24	ABN93373
26	39	5.1	4342	22	AAH54625
27	39	5.1	5126	24	ABL70493
28	39	5.1	13007	24	ABL32716
29	39	5.1	22073	22	AA104663
30	39	5.1	22073	23	ABL97570
31	39	5.1	1163020	24	AB067197
32	38.8	5.1	1005	14	AAQ40697
33	38.8	5.1	7108	24	ABK39966
34	38.8	5.1	2944528	24	ABA03041
35	38.4	5.0	343	23	AA550492
36	38.4	5.0	1021	18	AAV74478
37	38.4	5.0	1170	24	ABN91558
38	38.4	5.0	1194	23	AA552152
39	38.4	5.0	1401	21	AAAD0096
40	38.4	5.0	1413	21	AAA93966
41	38.4	5.0	1413	22	AAA91088
42	38.4	5.0	2538	13	AAQ27509
43	38.4	5.0	2700	21	AAAF0220
44	38.2	5.0	578	22	AAAF7743
45	38.2	5.0	630	21	AAAC48182

ALIGNMENTS

RESULT 1					
AAO71401	ID	AAO71401 standard; cDNA: 761 BP.			
AAO71401:	AC	AAO71401:			
01-APR-1995 (first entry)	XX				
House dust mite allergen DerfVII CDNA.	XX				
DerfVII allergen; antiallergic; allergy diagnosis; ss.	KW				
Dermatophagoides farinae.	XX				
Key	OS				
CDS	FT	location/Qualifiers			
misc_feature	FT	43..681			
misc_feature	FT	/*tag= a			
misc_feature	FT	94..111			
misc_feature	FT	/*tag= b			
misc_feature	FT	/note= "Df1 primer"			
misc_feature	FT	247..263			
misc_feature	FT	/*tag= c			
misc_feature	FT	/note= "Df4 primer"			
misc_feature	FT	559..582			
misc_feature	FT	/*tag= d			
misc_feature	FT	/note= "Df5 primer"			
misc_feature	FT	684..664			
misc_feature	FT	/*tag= e			
misc_feature	FT	/note= "Df6 primer"			
misc_feature	FT	225..208			
misc_feature	FT	/*tag= f			
misc_feature	FT	/note= "Df8 primer"			

Consensus sequence
Enterococcus faec
Plasmodium falci
Human anglogenis
Human immune sys
Genomic sequence
Long terminal repe
Streptococcus pne
Staphylococcus epi
S. epidermidis ope
S. epidermidis ope
S. epidermidis gen
S. epidermidis gen
Staphylococcus epi
S. epidermidis gen
Chemically treated
Human immune sys
Human reproductive
Human testicular a
Listeria innocua c
B. burgdorferi str
Human chemically p
Listeria monocytog
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Human zsl949 degen
Staphylococcus aur
S. aureus ftsa cod
Encodes alpha-acet
Plasmodium falci
Corynebacterium gl
Arabidopsis thalia

XX	WO9420614-A.
PN	
XX	15-SEP-1994.
PD	
XX	
FE	11-MAR-1994; 94WO-AU00117.
XX	
PR	12-MAR-1993; 93US-0031141.
PR	22-JUN-1993; 93US-0081540.
XX	
PA	(CHIL-) INST CHILD HEALTH RES.
XX	
PI	Chua K. Thomas WR;
XX	
DR	WPI; 1994-303021/37.
DR	P-PSDB; AAR60576.
XX	
PT	New nucleic acid encoding specific dust mite allergens - and
PT	related vectors, transformed cells, peptides and antibodies,
PT	useful for desensitisation and diagnosis.
XX	
PS	Claim 13; Page 38-39; 67pp; English.
XX	
CC	DerIVII antigen is useful as anti-allergic reagent for treating
CC	sensitivity to house dust mite allergens. The DNA can be used
CC	as a probe to detect the sensitivity of an individual to the
CC	allergen.
XX	
SO	Sequence 761 BP; 245 A; 134 C; 137 G; 245 T; 0 other;

Query Match	100.0%	Score 761:	DB 15	Length 761:
Best Local Similarity	100.0%	Pred. No.	1.5e-176:	
Matches 761: Conservative	0:	Mismatches	0:	Gaps 0:

Oy	1	GATCTATATCAATACATATCCAAAAAACATATCTTACAAATGATGAATTTTGTG	60
Db	1	GATCTTATATCAATTAACATCCAAAAAACATATCTTACAAAGATGAATTTTGTG	60
Oy	61	ATTGCTGCCGTGGCATTTGTGGCGTTGGCATGTCATATCTATGATGATGAATAC	120
Db	61	ATTGCTGCCGTGGCATTTGTGGCGTTGGCATGTCATATCTATGATGATGAATAC	120
Oy	121	GAGAAATCAACAAGCTATTGATGATGCCATTTGCTCTATTGAAACATCCGAAACATA	180
Db	121	GAGAAATCAACAAGCTATTGATGATGCCATTTGCTCTATTGAAACATCCGAAACATA	180
Oy	181	GATCCAAATGAATACCTGATCATGCCGATTAATTCGACAGCTCATTTGGTATTTGGAT	240
Db	181	GATCCAAATGAATACCTGATCATGCCGATTAATTCGACAGCTCATTTGGTATTTGGAT	240
Oy	241	TTCAAGGTGAATTAAGCCATCCGAACATTGAGGCTGAGAGTTGAACAATGAACGT	300
Db	241	TTCAAGGTGAATTAAGCCATCCGAACATTGAGGCTGAGAGTTGAACAATGAACGT	300
Oy	301	CAAGGTGATGCTAATGTCAAAAGGTGAAGAGGTATTGTTAAAGCTCATTTGTTGATCGT	360
Db	301	CAAGGTGATGCTAATGTCAAAAGGTGAAGAGGTATTGTTAAAGCTCATTTGTTGATCGT	360
Oy	361	GTTACAGTATATGCTGATGGATATGATTTTACATACAAATTTGGTGATCTTCAT	420
Db	361	GTTACAGTATATGCTGATGGATATGATTTTACATACAAATTTGGTGATCTTCAT	420
Oy	421	CCAACCATCATGTTCATTTGCGATTCAGATTTTGTGTGGCTGTGCCCTTGAAT	480
Db	421	CCAACCATCATGTTCATTTGCGATTCAGATTTTGTGTGGCTGTGCCCTTGAAT	480
Oy	481	TCTGATGAAGTAAACATTAACATGACATCTTTGAGTAGACAAATTCGTAATGTTGC	540
Db	481	TCTGATGAAGTAAACATTAACATGACATCTTTGAGTAGACAAATTCGTAATGTTGC	540
Oy	541	AACCATTTGGTGGCTTTCAATCTTGGATCCAAATTTTGGCGTTTATCTGATGATTG	600
Db	541	AACCATTTGGTGGCTTTCAATCTTGGATCCAAATTTTGGCGTTTATCTGATGATTG	600

Qy	601	ACCGGTAATTTTCCAGACACCCGTCACGTAAAGAAATGACCAAGATTTGGACACGACATTT	660
Db	601	ACCGGTAATTTTCCAGACACCCGTCACGTAAAGAAATGACCAAGATTTGGACACGACATTT	660
Qy	661	AAACGTGAATTGSGAAAAAATTAACCAATGACATCATTTTCCACCTGACAACTCTCTA	720
Db	661	AAACGTGAATTGSGAAAAAATTAACCAATGACATCATTTTCCACCTGACAACTCTCTA	720
Qy	721	TTTCACGTGACAAATAAATATTTTATTTTATTTTCTTCTCC	761
Db	721	TTTCACGTGACAAATAAATATTTTATTTTATTTTCTTCTCC	761

XX	RESULT 2
XX	AA071400
XX	ID AA071400 standard; CDNA; 812 BP.
XX	AA071400;
XX	01-APR-1995 (first entry)
XX	House dust mite allergen DerpVII CDNA.
DE	DerpVII allergen; anti-allergic; allergy diagnosis; ss.
XX	
XX	
OS	Dermatophagoides pteronyssinus.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	68..712
FT	/tag= a
FT	119..136
FT	misc_feature
FT	/tag= b
FT	/label= primer
FT	/note= "SEQ ID NO:3"
FT	272..288
FT	misc_feature
FT	/tag= c
FT	/label= primer
FT	/note= "SEQ ID NO:4"
FT	584..607
FT	misc_feature
FT	/tag= d
FT	/label= primer
FT	/note= "SEQ ID NO:5"
XX	
XX	
PN	W09420614-A.
XX	
PD	15-SEP-1994.
XX	
PF	11-MAR-1994; 94WO-AU00117.
XX	
PR	12-MAR-1993; 93US-0031141.
PR	22-JUN-1993; 93US-0081540.
XX	
PA	(CHIL-) INST CHILD HEALTH RES.
XX	
PI	Chua K, Thomas WR;
XX	
DR	WPI: 1994-303021/37.
DR	P-PSDB; AAR60575.
XX	
XX	New nucleic acid encoding specific dust mite allergens - and
PT	related vectors, transformed cells, peptides and antibodies,
PT	useful for desensitisation and diagnosis.
XX	
XX	Claim 3; Page 35-36; 67pp; English.
PS	
CC	DerpVII antigen is useful as anti-allergic reagent for treating
CC	sensitivity to house dust mite allergens. The DNA can be used
CC	as a probe to detect the sensitivity of an individual to the
CC	allergen.
XX	
XX	Sequence 812 BP; 270 A; 133 C; 142 G; 267 T; 0 other;

OS	Homo sapiens.
XX	MO200168912-A2.
XX	
XX	20-SEP-2001.
XX	
XX	15-MAR-2001; 2001WO-EP02955.
XX	
XX	15-MAR-2000; 2000DE-1013847.
XX	06-APR-2000; 2000DE-1019058.
XX	07-APR-2000; 2000DE-1019173.
XX	30-JUN-2000; 2000DE-1032529.
XX	01-SEP-2000; 2000DE-1043826.
XX	
XX	(EPIC-) EPIGENOMICS AG.
XX	
XX	Olek A, Piepenbrock C, Berlin K;
XX	WPI; 2001-602752/68.
XX	
XX	Claim 1; SEQ ID NO 289; 27pp; English.
XX	
XX	The invention relates to a nucleic acid comprising a sequence of 18
XX	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX	bisulphite, of genes associated with tumour suppression and
XX	oncogenes having a sequence taken from 536 (actually 533 since
XX	numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX	(Ss) and sequences complementary to (Ss). The nucleic acid may be a
XX	peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX	form part of a set of probes for detecting the cytosine methylation state
XX	and/or single nucleotide polymorphisms and also to be used in an
XX	array for analysing diseases associated with CpG dinucleotides e.g.
XX	cancers and tumours. The probes can also be used in a method for
XX	ascertaining genetic and/or epigenetic parameters for the diagnosis
XX	and/or therapy of existing diseases or the predisposition to specific
XX	diseases, by analysing cytosine methylations. The parameters may be
XX	compared to another set of genetic and/or epigenetic parameters, the
XX	differences serving as basis for diagnosis and/or prognosis events which
XX	are disadvantageous to patients. The present sequence is one of the
XX	533 genomic sequences derived from tumour suppressor genes and
XX	oncogenes.
XX	Note: The sequence data for this patent did not form part
XX	of the printed specification, but was obtained in electronic
XX	format directly from WIPO at
XX	ftp.wipo.int/pub/published_pcl_sequences.
XX	
XX	Sequence 7309 BP; 1851 A; 181 C; 1913 G; 3364 T; 0 other;
XX	
XX	Query Match 5.5%; Score 42; DB 22; Length 7309;
XX	Best Local Similarity 50.0%; Pred. No. 1.9;
XX	Matches 134; Conservative 0; Mismatches 130; Indels 4; Gaps 1
YY	493 AACATACATGACATCTTTTGAAGTAGACAAATTCGTAATGTTGTCAACATATTGGT 552
YY	2784 AATCTATCTAACCTATCTACACACCTCAATTAATACTCAAAATTCCTCTTCTTAACAT 2725
YY	553 GGTCTTTTCATCTTGGAATCCAAATTTTGGCGTTTATTCGATGT----ATTGACCCCTAT 608
YY	2724 TTTCTTAAAAAAATAATATTCATATTCACCCCTCTTTATATCAAAATTTAAAAATCTAAAT 2665
YY	609 TTTCGAAGACACCGTAGCTAGAGAAATGACAAAGTATTTGGACACGATTTAAACGTGA 668
YY	2664 CTATATAAAAAAATTCGTAAAAAAACCTCAAAAACCTACTGCTCTTATTTTAAATATA 2605
YY	669 ATTGGAAAAAATTAAACCAATGACATCTTTTCCAACTGTACATCTCTATTTCAGTG 728
YY	2604 AAAAAAATTAATTCAAAAAAATTAAAAAAAGCTTACCATTTTCAAAAGCAATCAATCTGCAA 2545

OY	729	ACCTAAATPAAATTTTATTATTATTTTAT	756
II			
Db	2544	ACTATAAATTAACCATCATTACT	2517
RESULT 5			
ID	ABK34010/c		
XX	ABK34010 standard; DNA: 7309 BP.		
AC	ABK34010:		
XX			
DT	18-JUN-2002 (first entry)		
XX			
DE	Human DNA for staging of Astrocytomas #49.		
XX			
KM	Human; de; astrocytoma; cytostatic; staging; cysteine methylation; CpG;		
KW	bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;		
XX	matrix assisted laser desorption/ionization mass spectrometry.		
OS	Homo sapiens.		
PN	WO200202808-A2.		
XX			
PD	10-JAN-2002.		
XX			
PF	02-JUL-2001; 2001WO-EP07538.		
XX			
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPIC-) EPIDENOMICS AG.		
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
DR	WPI; 2002-171649/22.		
XX			
PT	Novel chemically modified genomic DNA sequences, useful in the		
PT	characterisation, classification, differentiation, grading, staging,		
PT	treatment and/or diagnosis of astrocytomas or predisposition to		
PT	astrocytomas -		
PS			
XX			
XX	Claim 1; SEQ ID No 97; 37pp; English.		
CC	The invention relates to a nucleic acid comprising a sequence (I) of at		
CC	least 18 bases in length of a segment of chemically pre-treated genomic		
CC	DNA which has any one of the sequences of (ABK33919-ABK34032) or its		
CC	complement. Also included are an oligonucleotide or peptide nucleic		
CC	acid (or set thereof) of at least 9 nucleotides which hybridises to (I),		
CC	primers for (I), probes for detecting cytosine methylation or single-		
CC	nucleotide polymorphisms (SNP) in (I), an array of oligomers		
CC	or peptide nucleic acids for analysing diseases associated with the		
CC	methylation states of the CpG dinucleotides of (I). The array is useful		
CC	for determining genetic and/or epigenetic parameters, classification,		
CC	differentiation, grading, staging, treatment and/or diagnosis of		
CC	astrocytomas, or the predisposition to astrocytomas by analysing cytosine		
CC	methylation. Involves obtaining a biological sample containing genomic		
CC	DNA, extracting the genomic DNA, converting cytosine bases which are		
CC	unmethylated at the 5'-position, in the genomic DNA sample, to uracil or		
CC	another base which is dissimilar to cytosine in terms of hybridisation		
CC	behaviour. By chemical treatment and amplifying chemically pre-treated		
CC	genomic DNA fragments using the array and a polymerase, where the		
CC	amplificates carry a detectable label. The method further involves		
CC	identifying methylation status of one or more cytosine positions, and		
CC	analysing methylation status of the cytosine positions by reference to		
CC	one or more data sets. The genomic DNA is chemically treated by using a		
CC	bisulphite, hydrogen sulphite or disulphite. The amplification		
CC	step amplifies DNA which is of particular interest in astrocytoma or		
CC	brain tissue, based on the specific genomic methylation status of brain		
CC	tissues, as opposed to background DNA. The amplificates carry a		
CC	fluorescent label or radionuclide. Optionally, the labels of the		
CC	amplificates are detachable molecule fragments having a typical mass		
CC	which are detected in a mass spectrometer. The fragments of chemically		
CC	pre-treated genomic DNA to be amplified, have a single positive or		

CC negative charge for a better detectability in the mass spectrometer.
 CC Preferably, the amplification fragments of the amplification are
 CC detected by matrix assisted laser desorption/ionization mass spectrometry
 CC (MALDI) or using electron spray mass spectrometry (ESI). The
 CC present sequence is one of the chemically pre-treated reference DNA
 CC samples of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 7309 BP; 1851 A; 181 C; 1913 G; 3364 T; 0 other;

Query Match 5.5%; Score 42; DB 24; Length 7309;
 Best Local Similarity 50.0%; Pred. No. 1.9; Mismatches 130; Indels 4; Gaps 1;
 Matches 134; Conservative 0; Mismatches 130; Indels 4; Gaps 1;

OY 493 AACATTAACATGACATCTTTTGAAGTACGACAAATTCGTAATGTTGTCACATATTTGTT 552
 DB 2784 AATCTATCTAACCTATCTACACACCTAAATAATATCTCTTCTCTAAACT 2725
 OY 553 GGTCTTTCATCTTGATCCATTTTGGCGTTTATCTGATGT---ATTGACCGCTAT 608
 DB 2724 TTTCTAAATAAATAATTAATTCATATCCACCTCTTATATCAAAATTAATAATCTAAAT 2665
 OY 609 TTTCCAGACACCGTACGTAAGAAATGACAAAGTATGGACACAGATTAAACGTGA 668
 DB 2664 CTATTAATAAATAATCGTAATAAAACCTCAAAACCTACTCTTATTTTAAATATA 2605
 OY 669 ATTGGAATAAATAATTAACCAATAGACATCTTTTCCAACTGTACATCTCTATTTCACGTG 728
 DB 2604 AAAAAAATAAATAATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2545
 OY 729 ACAATAAATAAATAATTTTATTTTATT 756
 DB 2544 ACTATAAATAATATACCACTACTTACT 2517

RESULT 6

ABL33816/c
 ID ABL33816 standard; DNA; 7309 BP.

XX ABL33816;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1789.

XX Human; immune system disease; cytosine methylation; antiaesthatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmo-logical;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation

PS Claim 1; SEQ ID NO 1789; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 7309 BP; 1851 A; 181 C; 1913 G; 3364 T; 0 other;

Query Match 5.5%; Score 42; DB 24; Length 7309;
 Best Local Similarity 50.0%; Pred. No. 1.9; Mismatches 130; Indels 4; Gaps 1;
 Matches 134; Conservative 0; Mismatches 130; Indels 4; Gaps 1;

OY 493 AACATTAACATGACATCTTTTGAAGTACGACAAATTCGTAATGTTGTCACATATTTGTT 552
 DB 2784 AATCTATCTAACCTATCTACACACCTAAATAATATCTCTTCTCTAAACT 2725
 OY 553 GGTCTTTCATCTTGATCCATTTTGGCGTTTATCTGATGT---ATTGACCGCTAT 608
 DB 2724 TTTCTAAATAAATAATTAATTCATATCCACCTCTTATATCAAAATTAATAATCTAAAT 2665
 OY 609 TTTCCAGACACCGTACGTAAGAAATGACAAAGTATGGACACAGATTAAACGTGA 668
 DB 2664 CTATTAATAAATAATCGTAATAAAACCTCAAAACCTACTCTTATTTTAAATATA 2605
 OY 669 ATTGGAATAAATAATTAACCAATAGACATCTTTTCCAACTGTACATCTCTATTTCACGTG 728
 DB 2604 AAAAAAATAAATAATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2545
 OY 729 ACAATAAATAAATAATTTTATTTTATT 756
 DB 2544 ACTATAAATAATATACCACTACTTACT 2517

RESULT 7

ABL32887/c
 ID ABL32887 standard; DNA; 16439 BP.

XX ABL32887;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 860.

XX Human; immune system disease; cytosine methylation; antiaesthatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmo-logical;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 860; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SO Sequence 16439 BP; 4817 A; 191 C; 3420 G; 8011 T; 0 other;
Query Match 5.5%; Score 41.8; DB 24; Length 16439;
Best Local Similarity 56.0%; Pred. No. 2.7;
Matches 79; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 621 CGTACGTAAGAAATGACCAAGATTTGGCACCATTTTAAAGGTGAAATAA 680
DB 13129 CCTACTACAAATACCAACCTTAATTAATTAATTAATTAATTAATTAACAA 13070
QY 681 TTACCAATGACATCATTTTCCACTGTACATCTCTATTTGACGACATAATAA 740
DB 13069 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATCT 13010
QY 741 AATTTTATTTTATTTCTGC 761
DB 13009 AATTTCTACTTCTTTTACTAC 12989
RESULT 8
ABLI9335/C
ID ABLI9335 standard; DNA; 1367 BP.
XX
AC ABLI9335;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 9478.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX
PS Claim 1; SEQ ID NO 9478; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SO Sequence 1367 BP; 430 A; 284 C; 317 G; 336 T; 0 other;
Query Match 5.5%; Score 41.6; DB 23; Length 1367;
Best Local Similarity 51.0%; Pred. No. 1.5;
Matches 98; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 361 GTTCAGATGATATCGCTCGATGGAATATGATTTAGCATATCAATTTGGGATCTTCAT 420
DB 1226 GTTCAGTACGAGTGTCTGTGTACCGCTTAATAGTTGACTAATAATTTGCTGATCTGTT 1167
QY 421 CCAACACATCATGTCATTTGGATATTCAGATTTTGTGCTTGCCTTGAATTT 480
DB 1166 TGTACATATTTTGTGTTGAGTTTCCAGCTAGTCTGTGCTATGTAATTTAAACG 1107
QY 481 TCTGATGAGTAAATCAATACATGACATCTTTGAGTAGACACATTTGCTAATGTTGTC 540
DB 1106 CTGACGACGATCTAGTACGCTGAGCTTCTTTGGACATAGAAATTCATAGCAATGAT 1047
QY 541 AACCATATTTGTT 552
DB 1046 ATATATATATGT 1035
RESULT 9
ABLI9334/C
ID ABLI9334 standard; DNA; 4958 BP.
XX
AC ABLI9334;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 9475.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 9475; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU10840-ABU16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pot_sequences](http://wipo.int/pub/published_pot_sequences).
XQ Sequence 4958 BP; 1507 A; 987 C; 982 G; 1482 T; 0 other;

Query Match	5.58;	Score 41.6;	DB 23;	Length 4958;
Best Local Similarly	51.0%;	Pred. No. 2.2;		
Matches 98; Conservative	0;	Mismatches 94;	Indels 0;	Gaps 0;

Oy	361	GTTACGATGATATTCCTCGATGGAAATAGATTTAGCATACAAATTGGGTGATCTTCAT	420
Db	3439	GTTTACTAGGAGTGCTCTGTACGCCTTAATAAGTTTGACTAAAAAATGGGTGATCTAGTT	3380
Oy	421	CCAACCACTCATGCAATTTGGGATTTTCAAGAATTTTGCTGTGCCCTTGCCCTGAANT	480
Db	3379	TGTAAACAATTAATTTGGTTTGGAGTTTTCCAGCTAGAGTGCTTGTCGCAATGTGACTTTTAAAACG	3320
Oy	481	TCTGATGAAGGTAACATTAACATAGACATCTTTTGAAGTAGCAGCAANTGCTATGTTGTC	540
Db	3319	CTGAGGCACTTACTAGAGCTCTGACCTCTTTGGACCACTAAGAAATGCATAGACATGAT	3260
Oy	541	AACCATATTTGGT	552
Db	3259	ATATATATATGTT	3248

RESULT 10	
AAS21133	
ID	AAS21133 standard; cDNA; 484 BP.
XX	
AC	AAS21133;
XX	
DT	21-MAY-2002 (first entry)
XX	
DE	Consensus sequence of Tm12.84 isoforms with Tm 13.17 with B1 and B2.
XX	
KW	Anti-freeze peptide; Tm 12.84; yellow mealworm beetle; THP; cryoinjury;
KW	thermal hysteresis protein; cryosurgery; cold climatization; ss;
KW	hypothermic cell preservation; de-icing formulation; transgenic plant;
KW	transgenic animal; His tag; Tm 13.17; B1; B2.

XX	Tenebrio molitor.
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	14...418
FT	/*tag= a
FT	/product= "Thermal hysteresis protein"
FT	14...67
FT	/*tag= b
FT	68...415
FT	/*tag= c
FT	20...22
FT	/*tag= d
FT	/note= "Encodes Leu"
FT	26...28
FT	/*tag= e
FT	/note= "Encodes Leu"
FT	29...31
FT	/*tag= f
FT	/note= "Encodes Cys"
FT	32...34
FT	/*tag= g
FT	/note= "Encodes Phe"

FT	unsure	35..37	h	"Encodes Ala"
FT		/tag-		
FT		/note-		
FT	unsure	38..40	i	"Encodes Phe"
FT		/tag-		
FT		/note-		
FT	unsure	41..43	j	"Encodes Ala"
FT		/tag-		
FT		/note-		
FT	unsure	44..46	k	"Encodes Val"
FT		/tag-		
FT		/note-		
FT	unsure	47..49	l	"Encodes Glu"
FT		/tag-		
FT		/note-		
FT	unsure	50..52	m	"Encodes Leu"
FT		/tag-		
FT		/note-		
FT	unsure	53..55	n	"Encodes Gly"
FT		/tag-		
FT		/note-		
FT	unsure	56..58	o	"Encodes Ala"
FT		/tag-		
FT		/note-		
FT	unsure	59..61	p	"Encodes Leu"
FT		/tag-		
FT		/note-		
FT	unsure	68..70	q	"Encodes Asp"
FT		/tag-		
FT		/note-		
FT	unsure	74..76	r	"Encodes Glu"
FT		/tag-		
FT		/note-		
FT	unsure	77..79	s	"Encodes Phe"
FT		/tag-		
FT		/note-		
FT	unsure	80..82	t	"Encodes Val"
FT		/tag-		
FT		/note-		
FT	unsure	83..85	u	"Encodes Glu"
FT		/tag-		
FT		/note-		
FT	unsure	86..88	v	"Encodes Leu"
FT		/tag-		
FT		/note-		
FT	unsure	89..91	w	"Encodes Ala"
FT		/tag-		
FT		/note-		
FT	unsure	92..94	x	"Encodes Val"
FT		/tag-		
FT		/note-		
FT	unsure	95..97	y	"Encodes Glu"
FT		/tag-		
FT		/note-		
FT	unsure	98..100	z	"Encodes Leu"
FT		/tag-		
FT		/note-		
FT	unsure	101..103	aa	"Encodes Gly"
FT		/tag-		
FT		/note-		
FT	unsure	107..109	ab	"Encodes Val"
FT		/tag-		
FT		/note-		
FT	unsure	110..112	ac	"Encodes Glu"
FT		/tag-		
FT		/note-		
FT	unsure	113..115	ad	"Encodes Glu"
FT		/tag-		
FT		/note-		
FT	unsure	116..118	ae	"Encodes Glu"
FT		/tag-		
FT		/note-		
FT	unsure	119..121		

FT	/*tag= af	122.124	/*tag= "Encodes Gln"
FT	/*tag= ag	125..127	/*tag= "Encodes Val"
FT	/*tag= ah	137..139	/*tag= "Encodes Ser"
FT	/*tag= ai	140..142	/*tag= "Encodes Gln"
FT	/*tag= aj	143..145	/*tag= "Encodes Glu"
FT	/*tag= ak	146..148	/*tag= "Encodes Thr"
FT	/*tag= al	149..151	/*tag= "Encodes Ile"
FT	/*tag= am	152..154	/*tag= "Encodes Asp"
FT	/*tag= an	155..157	/*tag= "Encodes Lys"
FT	/*tag= ao	161..163	/*tag= "Encodes Val"
FT	/*tag= ap	167..169	/*tag= "Encodes Thr"
FT	/*tag= aq	170..172	/*tag= "Encodes Val"
FT	/*tag= ar	173..175	/*tag= "Encodes Leu"
FT	/*tag= as	188..190	/*tag= "Encodes Val"
FT	/*tag= at	194..196	/*tag= "Encodes Met"
FT	/*tag= au	197..199	/*tag= "Encodes Lys"
FT	/*tag= av	200..202	/*tag= "Encodes His"
FT	/*tag= aw	203..205	/*tag= "Encodes Val"
FT	/*tag= ax	209..211	/*tag= "Encodes Leu"
FT	/*tag= ay	212..214	/*tag= "Encodes Phe"
FT	/*tag= az	215..217	/*tag= "Encodes Ser"
FT	/*tag= ba	218..220	/*tag= "Encodes Lys"
FT	/*tag= bb	221..223	/*tag= "Encodes Lys"
FT	/*tag= bc	224..226	/*tag= "Encodes Thr"
FT	/*tag= bd		

FT	/note= "Encodes Gly"	227..229
FT	/tag= be	
FT	/note= "Encodes Val"	230..232
FT	/tag= bf	
FT	/note= "Encodes Leu"	233..235
FT	/tag= bg	
FT	/note= "Encodes Ala"	239..241
FT	/tag= bh	
FT	/note= "Encodes Leu"	245..247
FT	/tag= bi	
FT	/note= "Encodes Asp"	248..250
FT	/tag= bj	
FT	/note= "Encodes Thr"	
Query Match		
Best Local Similarity 16.9%; Pred. No. 1.6;		
Matches 77; Conservative 69; Mismatches 310; Indels 0; Gaps 0;		
QY	52 TTTTGTGATGGTCGCCGTGGCATTTGTGCGCGGTTGGCGTGATCCATTCACATATGAT	111
DB	26 TNNGTYYTNRYYTNYCYRYTNTNTTCNNAGTCGCGCTNACYNANGNANNTN	85
QY	112 AAAATCACCGAAGAATCAACAAGCTATGTGATGATGCATTCGTGCTATTCGAACATCC	171
DB	86 NAGNANNNGNCCNAAYCAGCRNNAARTGYNARRNNGNRRNRYGGAGTGCNNAGANRYN	145
QY	172 GAACAATAGATCCAAATGAAGTACCTGATCATGCGCATTAATTGGAACGTGATGTGT	231
DB	146 ATNRNNAAGYCGCANNNGTGCTNGGNGAYGAYCCYAAANTGAARRNNCANNTYTY	205
QY	222 ATTGTGATTTCAAGGTGAATTACCCATGCGAAACATTGAGCGTCGAGATTTGAACAA	291
DB	206 TGCNTNNYNARRRNNYRGYRNTRGYNRCNGAANCNGAGANRYRNRNGYANRYRYTN	265
QY	292 ATGAACGTCGAAGTGATGCTAATGTCAAAGGTGAAGGGATGTAAAGCTCATTTG	351
DB	266 AARGNNAAGTGANRRNNNNNNNNNNRRARRRRYRRARRRNTNNYNNRRARN	325
QY	352 TTGATCGGTGTTCAGATGATATCGTCTCGATGAATGATTTAGCATACAAATTGGGT	411
DB	326 NNNNNNNNNNNGARRRRNNNNNNRRRRNNNNNTNRRANNYYAANNNNNNNNN	385
QY	412 GATCTTCATCCAAACCTCATGCTATTTGGATATTCAGATTTTGTGTCCTTGCC	471
DB	386 RRNANNARRCCRRNTTYTYNCNRRYTRNTTENNNTNNNNNNNNYNGNRRNRRTYBAN	445
QY	472 CTTGAATTTCGATGAAGGTACATACATGACA	507
DB	446 AATAAAGNNNTTNTNRTNNRRMAAAAAA	481
RESULT 11		
AAK13108		
ID	AAK13108 standard; DNA; 11871 BP.	
AC	AAK13108;	
XX		
DT	19-MAR-1999 (first entry)	
XX		
DE	Enterococcus faecalis genome contig SEQ ID NO:171.	
XX		
KW	Enterococcus faecalis; contig; detection: Enterococcal infection;	
KW	vaccine; attenuation; computer readable medium, ds.	
XX		
OS	Enterococcus faecalis.	
XX		
PN	W09850555-A2.	

[illegible]

Pd				11-MAY--2000.	
Xx					
Pf	XX	PF	05-NOV-1999;	99WO-US26796.	
PX					
Pr	xx	PR	05-NOV-1998;	98US-0107131.	
xx					
Pa	(HOFF/)	PA	HOFEMAN S.		
pa	(CARU/)	PA	CARUCI D.		
PA	(GARDN/)	PA	GARDNER M.		
pA	(VENT/)	PA	VENTER J C.		
Pi					
xP	Hoffman S,				
PI	Carucci D,				
xx	Venter J,				
Dx	WPI:				
xx	2000-365347/31.				
PT					
pt	Proteins encoded by chromosome 2 of the human malarial parasite,				
Pt	Plasmodium falciparum, useful as antimalarial vaccines and in the				
px	diagnosis of P.falciparum infection -				
PS					
ps	Disclosure; Page 565-566; 577pp; English.				
xx					
CC	The present invention describes proteins and their fragments (I) encoded				
cc	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.				
CC	Also described are: (1) nucleotide sequences (II) encoding (I); and (2)				
cc	vaccines against P. falciparum infection comprising (I) or (II).				
CC	(I) and (II) are useful for the development of vaccines against				
cc	P. falciparum infection. (I) and polyclonal antisera or a monoclonal				
CC	antibody raised to immunogens comprising the sequences of (I), are				
cc	useful in the detection of infection with P. falciparum. Furthermore,				
CC	(I) (especially when they are rifins or secreted or membrane proteins)				
cc	can aid the identification of drugs to treat or prevent P. falciparum				
CC	infection, or they can be used to identify drug resistance in				
cc	P. falciparum. Sequencing of the plasmodium chromosome 2 and the				
CC	subsequent identification of proteins encoded by it will help to expand				
cc	our understanding of parasite biology, a process hampered by the				
CC	complexity of the parasitic lifecycle, and provide new targets for				
cc	vaccine and drug development. Parasite resistance to drugs and mosquito				
CC	resistance to insecticides have led to a resurgence of malaria in many				
cc	parts of the world, and there is a pressing need for vaccines and new				
CC	drugs. AAT00078 to AAT00287 and AAB18144 to AAB18352 represent nucleotide				
cc	and protein sequences given in the present invention, but which are not				
CC	specifically mentioned within the specification.				
Xx					
SQ	Sequence 4677 BP; 2106 A; 402 C; 966 G; 1203 T; 0 other:				
	Query Match	5.3%; Score 40.2; DB 21; Length 4677;			
	Best Local Similarity	43.2%; Pred. No. 4.7;			
Matches	189; Conservative	0; Mismatches	248; Indels	0; Gaps	0
Oy	23	AAAAAAAACATTCCTTACAAATAGATGAATAATTGGTTGATTCGTGCCCGGCATTGGTCG	82		
Db	3830	AGAAAAAAAATTACA AAAAAGTAAGTGTTGATTTTAAGAAGTAAGAACCAAAAGATG	3889		
Oy	83	CGTTTTGGCGTCATCCAAATTCATGATGATGAATAATCACCGAAGAATCACA CAAGGATTTG	142		
Db	3890	AATATGTTACAAGTTCAAATGAAGTAGAAGATATACATGTCGAAGTTATRGAAGCAAGATGAG	3949		
Oy	143	ATGATGCCATTCTGCTGTTTGAACAATCCGAAACCAATAGATCCAAATGAAAGACTGATC	202		
Db	3950	AAGAAAGATATTAGAAAGAAAGATTAAGTTGAAGATATACATGAAATATAGATGAAGATATAG	4009		
Oy	203	ATGCCGATTAATTCGAAAGCTCATGTGGTATTGTGGATTTCAAAGGTGAATTCACCATGC	262		
Db	4010	ATGAAGATATTAGCTGAAGCAAAAGATGAAGTTATTAATTTAAATAGTCCA AAAGAGAAC	4069		
Oy	263	GAAACATTATAGGCTCGACGATTCGAACCAATGAAAGCTCAAGGTGATGCTAAATGTCAAAG	322		
Db	4070	GCATTGAAAAGGTTAAAGAGAAAAAGAAAAAATTTAGAAAAAAGTTGAAGAAGGCTTTA	4129		
Oy	323	GTGAAGAGGCTATTGTTAAAGCTCATTTGTGTGATGCGGTGTCAGATGATATCGTCTCGA	382		
Db	4130	GTGCTCTTTAAAAAACACAGTACGAAGATTAATGCAAAATATGTTCAAAAATTTGATRAAGAG	4189		

QY 383 TCGAATATGATTGACATACAAATGGGATCTTCATCCACACTCATGCTATTTGG 442
DB 4190 TTGATTAAGAGATCTTAAGCTTTAGAAATCAAAAATGATGTACTAATGTTTAAAC 4249
QY 443 ATATTCAGATTTGTT 459
DB 4250 AAAATCAAGATTTTTTT 4266

RESULT 13
AB067097/c
ID AB067097 standard; DNA: 14429 BP.

AC AB067097;

DT 28-AUG-2002 (first entry)

DE Human angiogenesis associated polynucleotide SEQ ID NO 127.

XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antilatheriosclerotic; ds.

OS Homo sapiens.

PN WO200246454-A2.

PD 13-JUN-2002.

PF 06-DEC-2001; 2001WO-EP14320.

PR 06-DEC-2000; 2000DE-1061338.

PA (EPIC-) EPIGENOMICS AG.

PI Schacht O;

DR WPI; 2002-500450/53.

PT New nucleic acid fragments from chemically treated
PT angiogenesis-associated genes, useful for determining methylation
PT status, e.g. in diagnosis or treatment of cancer -
PS Claim 1; SEQ ID NO 127; 41pp + Sequence Listing; German.

XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (AB066971-AB067178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.

XX Sequence 14429 BP; 3457 A; 325 C; 3602 G; 7045 T; 0 other;

Query Match 5.3%; Score 40; DB 24; Length 14429;

Best Local Similarity 53.9%; Pred. No. 7.2; Mismatches 70; Indels 0; Gaps 0;

Matches 82; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 606 TATTTTCCAGACCGTAGTAAGAAATGACCAAGTATTTGGACACGATTTAAACG 665

DB 11868 TTTCTTCCAACTATCATCATATCTACCTCCCAAAATTAACACATCTTAATAAAT 11809

QY 666 TGAATTTGAAAAAATTAACCAATAGACATATTTTCCACAGCTACATCTCTATTTC 725

DB 11808 TTAATAATTTCAAAAAAATAAAAAATCATTTTCTATTACTATAATAATTT 11749
QY 726 CTGACATATAAATAATTTTATTATTATT 757
DB 11748 CCCATAATTTAAAAAACAATCAAAATCATAT 11717

RESULT 14
ABL34242/c
ID ABL34242 standard; DNA: 14429 BP.

AC ABL34242;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2215.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antilatheriosclerotic; antihaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; ARDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
PS Claim 1; SEQ ID NO 2215; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX Sequence 14429 BP; 3457 A; 325 C; 3602 G; 7045 T; 0 other;

Query Match 5.3%; Score 40; DB 24; Length 14429;

Best Local Similarity 53.9%; Pred. No. 7.2; Mismatches 70; Indels 0; Gaps 0;

Matches 82; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 606 TATTTTCCAGACCGTAGTAAGAAATGACCAAGTATTTGGACACGATTTAAACG 665

DB 11868 TTTCTTCCAACTATCATCATATCTACCTCCCAAAATTAACACATCTTAATAAAT 11809

QY 666 TGAATTTGAAAAAATTAACCAATAGACATATTTTCCACAGCTACATCTCTATTTC 725

DB 11808 TTAATAATTTCAAAAAAATAAAAAATCATTTTCTATTACTATAATAATTT 11749

QY 726 CTGACATAAATAAATTTTATTATT 757

Db 11748 CCCATTAATTAATAAACATCAAAATCATAT 11717

RESULT 15

ABA90521/c
ID ABA90521 standard; DNA; 2365589 BP.

AC ABA90521;

DT 16-MAY-2002 (first entry)

DE Genomic sequence of *Lactococcus lactis* IL1403.

KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.

OS *Lactococcus lactis* IL1403.

PN FR2807446-A1.

PD 12-OCT-2001.

PF 11-APR-2000; 2000FR-0004630.

PR 11-APR-2000; 2000FR-0004630.

PA (INRG) INRA INST NAT RECH AGRONOMIQUE.

PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;

DR WPI: 2002-043418/06.

PT New nucleotide sequence useful in the identification or *Lactococcus*
lactis and related species -

PS Claim 1; SEQ ID 1; 2504pp; French.

CC The present invention is related to a *Lactococcus lactis* nucleotide
CC sequence (ABA90521) and related proteins (AB853300-AB855621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.

CC Note: The sequence data for this patent is based on equivalent patent
CC WO20017734 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

Query Match 5.3%; Score 40; DB 24; Length 2365589;

Best Local Similarity 44.9%; Pred. No. 29;

Matches 151; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

YY 9 ATCATATAACATCAAAATATATCTTACAAATGATGAATTTTGTGATTGCTGC 68

Db 2324719 ATCAATTTACATGAAATTCATCCCAACAGACGTATGAGCTTGCTGTA 2324660

YY 69 CGTGCAATTTGCGCCGTTTGGCTGATCCAAATTCATATGATTAATCACCAGAAAT 128

Db 2324659 AGATGCAAGATGACCTTTTATACGAACTCAAGATTCATGAAACGGAACATG 2324660

YY 129 CAACAAAGCTATGATGATGCTGCTGATTTGAAACATTCGAAACATAGATCAAT 188

Db 2324599 TCGCGGGGAGACGCTTTCATGACCAATGGCAGCCCTTTTGAAGTTGATGCAA 2324540

YY 189 GAAGTACCTGATGATGCGGATTAATTCGAGCTCATTTGATTTGATTTCAAGG 248

Db 2324539 TGGTGTCAATGAGCTTACGTAACAACTCTATCCGATACATCAGCTGTGG 2324480

YY 249 TGAATTGACCTGCGAAACATTTGAGGCTCGAGATTTGAACAAATGAAACGTCAGGTGA 308

Db 2324479 CGTTTTGCCAGACAGACGCTTCACCGCTGTTAAATGAAGAGTGGCTGTAAGATAT 2324420

YY 309 TGCATATGTCAAAGGTGAGAGGCTATGTTAAAGC 344

Db 2324419 TCGTCTAGTATTTTTCACCGCTGTGCAATCAAC 2324384

RESULT 16

AAV52500/c
ID AAV52500 standard; DNA; 1702 BP.

AC AAV52500;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:367.

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

PN MO9818931-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; 97MO-US19588.

PR 31-OCT-1996; 96US-0029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

DR Kunsch CA, Rosen CA;

DR WPI: 1998-27225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
pneumoniae

PS Claim 1; Page 1382-1383; 1409pp; English.

CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridize to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX SQ Sequence 1702 BP; 422 A; 551 C; 398 G; 331 T; 0 other;

Query Match 5.2%; Score 39.6; DB 19; Length 1702;

Best Local Similarity 43.3%; Pred. No. 5;

Matches 186; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the *S. pneumoniae* genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence an
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the *S. pneumoniae* genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the *S. pneumoniae* genome of commercial importance, or
CC expression modulating fragments of the *S. pneumoniae* genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for *S. pneumoniae*.
XX
SQ Sequence 32768 BP, 9834 A; 6168 C 7154 G; 9609 T; 3 other;

Query Match	5.2%	Score 39.6	DB 19	Length 32768
Best Local Similarity	43.3%	Pred. No. 11		
Matches 186; Conservative	0	Mismatches 244	Indels 0	Gaps 0

Qy	63	TGCTGCCGTGGCAATTGGTTCGCCGGTTTGGCGTGAATCCAAATTCAGATGATTAATAATCAACCGA	1222
Db	541	TGAAGCGGAGCACTCGTTGATGCTGAGGCTGAGGCCCTAGTACTTGCTGAGGCTGACGC	4828
Qy	123	AGAAATCAACAAAGCTATTGATGATGSCAATTGCTGCTATTGAAACAATCCGAACAATAGA	1822
Db	481	ACTGGTTGATGACCCTATTCAGATGCACATCGTTGTGAGGCCGACGACACTTGTGACGCTGAGGC	4222
Qy	183	TCCAAATGAAGTACCTCATCATCCTGCGAATATTCGAAGCGTACATGTTGATTTGAGATT	2424
Db	421	TGAAGCACTGTAATTCTGCTGAGGCTTGACGCACCTGTAATCTTGCTGAACCTGAGGGCTTGT	3622
Qy	243	CAAAAGTGAATTTAGCCATTCGCAAAACATTTGAGGCTCGAGGATTTGAACAATGAACCTCA	3020
Db	361	CGATGCTGAGGCTGACGCACACTGGTTGATGACAGATTGAGAAGCTGATGTGCTTGCTGAGGC	3022
Qy	303	AGGTGATGCTAATGTCAAAGGTGAAAGGGGATTTGTTAAAGCTCATTTTGTGATCGGTGT	3622
Db	301	TGACGTACTCGTTGATGCTCTGACTGAGCGGACATCGTTGAGCGTGAAAGCCGAACACATTTGT	2422
Qy	363	TCACGATGATATCGTCTCGATGGAATATGATTTAGCATACAAATTTGGGTGATCTTATCC	4222
Db	241	TGACGCTGATGCTGAGGCGCTAGTACTTGCTGAAGCGGAGCACACTCGTTGATGCCGATTC	1822
Qy	423	AACCACTCATGTCAATTTCCGATATTCAGAATTTTGTGTTGCTTGCCCTGACCTTGAAATTTTC	4822
Db	181	AGACGCTGAGATTAATCTCTGAGGCTGAAGCACATTTGCTCTTGTAAGAACGATGCACATTTGT	1222
Qy	483	TGATGAAGGT 492	
Db	121	TGATGCTGAT 112	

XX	30-APR-2002.
PD	
XX	13-AUG-1998;
FE	98US-0134001.
XX	14-AUG-1997;
PR	97US-055779P.
PR	08-NOV-1997;
XX	97US-064964P.
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Doucette-Stamm LA, Bush D;
XX	
DR	WPI; 2002-381255/41.
DR	P-PSDB; ABP38519.
XX	
PT	Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX	polypeptide, useful for diagnosing and treating bacterial infections -
XX	Disclosure; SEQ ID 527; 267bp; English.
XX	

Query Match	5.2%	Score	39.4	DB	24	Length	276
Best Local Similarity	50.3%	Pred. No.	3.4				
Matches	97	Conservative	0	Mismatches	96	Indels	0
						Gaps	0

QY	566	TGATCAACATTTTGGCGTTTAATCTCATGTAAATGACCAGCGATTTCACAGACACCGTAC	625
Db	240	TTTGATCAATGTATGATACACTTTTCTCAAGTATTGCCTCTTTTGGAAAAAGAACGCTAC	181
QY	626	GTAAGAATAATGACCAAGATTTGGCACACGACATTTAATCGTAATGGAAAAAATTTAAC	685
Db	180	ATAACCTTTAAATCCATGATATTACACAGACATTTTATFTGTTATCTGATATAACAAGATG	121
QY	686	CAATAGACATCATTTTTCCAACTGTACATCTCATTTTCACTGCAATTAATAATAAAATT	745
Db	120	CATCCATTTTAAATATTTCTTCACAGTCACTTCAATTAATAATTTCTCAAAAAATTT	61
QY	746	TTATTTTATTTTC	758
Db	60	ATAACAATCTTGC	48

XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Kimerly MJ;
XX
XX WPI; 2001-316495/33.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 1851-1852; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 3238 BP; 1095 A; 520 C; 427 G; 1196 T; 0 other;
SQ

Query Match 5.2%; Score 39.2; DB 22; Length 3238;
Best Local Similarity 57.3%; Pred. No. 7.5;
Matches 71; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 122 AAGAAATCAACAAGCTATTGATGATCCATTGCTCTATTGGAACAATCCGAACAATAG 181
II IIII II IIII IIII IIII IIII IIII IIII IIII
DB 2724 AACAAATCCCTAATATTATTCATGACATCTTGATGATGAACAACCAATCCTTAG 2783
OY 182 ATCCAAATGAAAGTACTGATCATGCCGATGAATTCGAACGTCATGTTGGATTGTGGATT 241
II IIII IIII IIII IIII IIII IIII IIII IIII IIII
DB 2784 ATAAAGTGAACAACATCTATTATTTCGCTTACACCTCAATCTAAATAAAGGTAGCTA 2843
OY 242 TCAA 245
IIII
DB 2844 TCAA 2847

RESULT 23
AAH54467
ID AAH54467 standard; DNA; 3931 BP.
XX
XX AAH54467;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3831.
DE Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX

OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Kimerly MJ;
XX
XX WPI; 2001-316495/33.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 1471-1472; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 3931 BP; 1452 A; 628 C; 567 G; 1284 T; 0 other;
SQ

Query Match 5.2%; Score 39.2; DB 22; Length 3931;
Best Local Similarity 57.3%; Pred. No. 7.9;
Matches 71; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 122 AAGAAATCAACAAGCTATTGATGATCCATTGCTCTATTGGAACAATCCGAACAATAG 181
II IIII II IIII IIII IIII IIII IIII IIII IIII
DB 2576 AACAAATCCCTAATATTATTCATGACATCTTGATGATGAACAACCAATCCTTAG 2635
OY 182 ATCCAAATGAAAGTACTGATCATGCCGATGAATTCGAACGTCATGTTGGATTGTGGATT 241
II IIII IIII IIII IIII IIII IIII IIII IIII IIII
DB 2636 ATAAAGTGAACAACATCTATTATTTCGCTTACACCTCAATCTAAATAAAGGTAGCTA 2695
OY 242 TCAA 245
IIII
DB 2696 TCAA 2699

RESULT 24
AAH54213
ID AAH54213 standard; DNA; 4241 BP.
XX
XX AAH54213;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3577.
DE Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX

```
XX OS Staphylococcus epidermidis.
XX
XX MO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX KImmery WJ;
XX
XX WPI; 2001-316495/33.
XX
XX
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX
XX PS Claim 8; Page 1164-1166; 2188bp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce host cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX CC N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX SQ Sequence 4241 BP; 1541 A; 676 C; 578 G; 1446 T; 0 other;
XX
XX Query Match 5.2%; Score 39.2; DB 22; Length 4241;
XX Best Local Similarity 57.3%; Pred. No. 8.1;
XX Matches 71; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
XX
XX QY 122 AAGAATCAACAAGCTATTGATGCGATTGCTGTATTGAACAATCCGAACAATAG 181
XX || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 3659 AACAAATCCCTAATATATATCATGACATCTTGATGATATGAACAAACCAATCCTTAG 3718
XX
XX QY 182 ATCCAATGAAGTACCTGATATGCGCATTAATTCGAACGTCATGTTGGTATGCGATT 241
XX || || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
XX Db 3719 ATAAAGTGAAACATCTTATATTCGCTTACACCTCAATCTAAATAAAGTAGCTA 3778
XX
XX QY 242 TCAA 245
XX |||||
XX Db 3779 TCAA 3782
XX
XX
XX RESULT 25
XX ABN93373
XX ID ABN93373 standard; DNA: 1350 BP.
XX
XX AC ABN93373;
XX
XX DT 24-JUL-2002 (first entry)
XX
XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2836.
XX
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
```

```
KW antibacterial; gene therapy; gene; ds.
XX
XX OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX PA Doucette-Stamm LA, Bush D;
XX
XX PI
XX
XX DR WPI; 2002-381255/41.
XX
XX DR P-PSDB; ABP40828.
XX
XX
XX PT Novel isolated nucleic acid encoding a staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
XX PS Disclosure; SEQ ID 2836; 267bp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP5124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
XX SQ Sequence 1350 BP; 474 A; 178 C; 252 G; 446 T; 0 other;
XX
XX Query Match 5.1%; Score 39; DB 24; Length 1350;
XX Best Local Similarity 56.7%; Pred. No. 6.6;
XX Matches 72; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
XX
XX QY 134 AAGCTATTGATGATGACATGCTGCTATTGAAACAATCCGAACAATAGATCAATGAAG 193
XX || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 1007 AATTTGTTGGCAATCAGACGAAATTTGATGATGACGAAGTTAAGCATATAGTTAAC 1066
XX
XX QY 194 TACCTGATCATGCCGATTAATTCGAACGTCATGTTGGTATTGCGATTTCAAAGCTAAT 253
XX || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
XX Db 1067 TTGATGATCAATCACTTATCAAAATTAATGCTGCTTTTACTTGATGATTTAAATGAAAAAT 1126
XX
XX QY 254 TAGCCAT 260
XX || ||||
XX Db 1127 TTGGCAT 1133
XX
XX
XX RESULT 26
XX AAH54625
XX ID AAH54625 standard; DNA: 4342 BP.
XX
XX AC AAH54625;
XX
XX DT 03-SEP-2001 (first entry)
XX
XX DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3989.
XX
XX KW Staphylococcus epidermidis sri strain; infection; diagnosis;
XX vaccination; endocarditis; ds.
XX
XX OS Staphylococcus epidermidis.
XX
XX MO200134809-A2.
XX
XX
```

XX	10-JAN-2002.
PD	
XX	
PF	29-JUN-2001; 2001WO-EP07471.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIC-) EPICENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI: 2002-154758/20.
XX	
PT	Nucleic acid, useful for diagnosis and therapy of diseases associated
PT	with cell signalling e.g. cancer, comprises chemically modified genomic
PT	sequences of genes associated with cell signalling
PS	
XX	Claim 1; SEQ ID NO 383; 24pp;sequence listing; English.
XX	
CC	The invention relates to a nucleic acid comprising a sequence of at least
CC	18 bases of a segment of chemically pretreated DNA of genes associated
CC	with cell signalling. The activity of the modified sequences of the
CC	invention may be described as cytostatic. The object of the invention is
CC	to provide the chemically modified DNA of genes associated with cell
CC	signalling, as well as oligonucleotides and/or PM-oligoners for
CC	detecting cytosine methylations, as well as a method which is
CC	particularly suitable for the diagnosis and/or therapy of genetic and
CC	epigenetic parameters of genes associated with cell signalling. The
CC	chemically modified DNA provided by the invention is useful for diagnosis
CC	and therapy of diseases such as solid tumours and cancer. The sequences
CC	given in records ABL70111-ABL70626 represent chemically pre-treated
CC	genomic DNA's of genes associated with cell signalling.
CC	Note: The sequence data for this patent is not represented in the printed
CC	specification, but is based on sequence information supplied by the
CC	European Patent Office.
XX	
XX	Sequence 5126 BP; 1633 A; 26 C; 1012 G; 2455 T; 0 other;
XX	
QY	Query Match 5.1%; Score 39; DB 24; Length 5126;
QY	Best Local Similarity 63.2%; Pred. NO. 9.5;
Db	Matches 60; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY	660 TAAACGTGAATTGGAATAAATTAACAATAGACATCATTTTTCGAACGTGACAAATCTCT 719
Db	4020 TAAATTAATACCTTAAAAAAGCTTTAAACAACAACCAACCCCTAATAATATCTACAAACCT 3961
QY	720 ATTTCACGTGACATTAATAATTAATTTTATTTT 754
Db	3960 AATTTTAAAAACATTTATTTATTAATAATCTCATATATTT 3926
XX	
XX	RESULT 28
XX	ABL32716/C
ID	ABL32716 standard; DNA; 12007 BP.
XX	
XX	ABL32716;
XX	
XX	26-MAR-2002 (first entry)
XX	
XX	Human immune system associated gene SEQ ID NO: 689.
XX	
XX	Human; immune system disease; cytosine methylation; antiasthmatic;
KW	antiartherosclerotic; antianemic; cytostatic; nootropic;
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW	antirheumatic; antiarthritic; antididiabetic; antiporiatic;
KW	antitumorigenic; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW	gene; ds.
XX	
XX	Homo sapiens.
XX	

PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 689; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 12007 BP; 3469 A; 114 C; 2785 G; 5639 T; 0 other;
Query Match 5.1%; Score 39; DB 24; Length 12007;
Best Local Similarity 56.7%; Pred. No. 12;
Matches 72; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 628 AAGCAATGACCAAGATGTGGACGACGATTTAAAGTGAATGAAAAATTACCA 687
DB 9860 AAATAATTAACACAAATCTACCTCAATAATTAATTAATTAATTAATTAATCC 9801
QY 688 ATGACATCATTTTTCACACTGACATCTCTATTTCACGACATTAATAATTAATTTT 747
DB 9800 AAAAAAATTTTATATATCTACCTACATACACATTAATAATTTCAATAATTAATTTT 9741
QY 748 ATTTTTA 754
DB 9740 TTTTATA 9734
RESULT 29
AAL04663/c
ID AAL04663 standard; DNA; 22073 BP.
XX
AC AAL04663;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 7351.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

PR	14-AUG-2000	2000US-0225267.
PR	14-AUG-2000	2000US-0225267.
PR	14-AUG-2000	2000US-0225268.
PR	14-AUG-2000	2000US-0225270.
PR	14-AUG-2000	2000US-0225447.
PR	14-AUG-2000	2000US-0225757.
PR	14-AUG-2000	2000US-0225758.
PR	14-AUG-2000	2000US-0225759.
PR	18-AUG-2000	2000US-0226271.
PR	22-AUG-2000	2000US-0226681.
PR	22-AUG-2000	2000US-0226868.
PR	22-AUG-2000	2000US-0227182.
PR	23-AUG-2000	2000US-0227009.
PR	30-AUG-2000	2000US-0228924.
PR	01-SEP-2000	2000US-0229287.
PR	01-SEP-2000	2000US-0229343.
PR	01-SEP-2000	2000US-0229344.
PR	01-SEP-2000	2000US-0229345.
PR	05-SEP-2000	2000US-0229509.
PR	05-SEP-2000	2000US-0229513.
PR	06-SEP-2000	2000US-0230437.
PR	06-SEP-2000	2000US-0230438.
PR	08-SEP-2000	2000US-0231242.
PR	08-SEP-2000	2000US-0231243.
PR	08-SEP-2000	2000US-0231244.
PR	08-SEP-2000	2000US-0231413.
PR	08-SEP-2000	2000US-0231414.
PR	08-SEP-2000	2000US-0232080.
PR	08-SEP-2000	2000US-0232081.
PR	12-SEP-2000	2000US-0231968.
PR	14-SEP-2000	2000US-0232397.
PR	14-SEP-2000	2000US-0232398.
PR	14-SEP-2000	2000US-0232399.
PR	14-SEP-2000	2000US-0232400.
PR	14-SEP-2000	2000US-0232401.
PR	14-SEP-2000	2000US-0233063.
PR	14-SEP-2000	2000US-0233064.
PR	14-SEP-2000	2000US-0233065.
PR	21-SEP-2000	2000US-0234223.
PR	21-SEP-2000	2000US-0234274.
PR	25-SEP-2000	2000US-0234997.
PR	25-SEP-2000	2000US-0234998.
PR	26-SEP-2000	2000US-0235484.
PR	27-SEP-2000	2000US-0235834.
PR	27-SEP-2000	2000US-0235836.
PR	29-SEP-2000	2000US-0236337.
PR	29-SEP-2000	2000US-0236367.
PR	29-SEP-2000	2000US-0236368.
PR	29-SEP-2000	2000US-0236369.
PR	29-SEP-2000	2000US-0236370.
PR	02-OCT-2000	2000US-0236802.
PR	02-OCT-2000	2000US-0237037.
PR	02-OCT-2000	2000US-0237038.
PR	02-OCT-2000	2000US-0237039.
PR	02-OCT-2000	2000US-0237040.
PR	13-OCT-2000	2000US-0239355.
PR	13-OCT-2000	2000US-0239397.
PR	20-OCT-2000	2000US-0240960.
PR	20-OCT-2000	2000US-0241221.
PR	20-OCT-2000	2000US-0241785.
PR	20-OCT-2000	2000US-0241786.
PR	20-OCT-2000	2000US-0241787.
PR	20-OCT-2000	2000US-0241808.
PR	20-OCT-2000	2000US-0241809.
PR	20-OCT-2000	2000US-0241826.
PR	01-NOV-2000	2000US-0244617.
PR	08-NOV-2000	2000US-0246474.
PR	08-NOV-2000	2000US-0246475.
PR	08-NOV-2000	2000US-0246476.
PR	08-NOV-2000	2000US-0246477.
PR	08-NOV-2000	2000US-0246478.
PR	08-NOV-2000	2000US-0246523.
PR	08-NOV-2000	2000US-0246524.
PR	08-NOV-2000	2000US-0246525.
PR	08-NOV-2000	2000US-0246526.
PR	08-NOV-2000	2000US-0246527.
PR	08-NOV-2000	2000US-0246528.
PR	08-NOV-2000	2000US-0246532.
PR	08-NOV-2000	2000US-0246569.
PR	08-NOV-2000	2000US-0246610.
PR	08-NOV-2000	2000US-0246611.
PR	08-NOV-2000	2000US-0246613.
PR	17-NOV-2000	2000US-0249207.
PR	17-NOV-2000	2000US-0249208.
PR	17-NOV-2000	2000US-0249209.
PR	17-NOV-2000	2000US-0249210.
PR	17-NOV-2000	2000US-0249211.
PR	17-NOV-2000	2000US-0249212.
PR	17-NOV-2000	2000US-0249213.
PR	17-NOV-2000	2000US-0249214.
PR	17-NOV-2000	2000US-0249215.
PR	17-NOV-2000	2000US-0249216.
PR	17-NOV-2000	2000US-0249217.
PR	17-NOV-2000	2000US-0249218.
PR	17-NOV-2000	2000US-0249244.
PR	17-NOV-2000	2000US-0249245.
PR	17-NOV-2000	2000US-

Db	15294	TCCAGGTCGAAGTCATGTGAAATTTGGTAATTTGGTAAGTTCGAAAGTCACCTTTTGGATTTA	15233			
QY	360	TGTTTCACGATCATATTCGTCCTCGATGAGTAATGATTTAGCATPACAAATGGGCGATCTTCA	419			
Db	15234	AAATACGCTGTAACTTACACCTGGGTTTATCAGGAGTGAATCACAAGTCGTGAGCATCT	15175			
OY	420	TCCAACACGTCATGTCATTTCCGGATAT	446			
Db	15174	GTAATTCGTCATTTAAATGATGATAT	15148			
RESULT 31						
ID	ABQ67197/c	ABQ67197/c				
XX	ABQ67197 standard; DNA: 1163020 BP.					
AC	ABQ67197:					
XX						
DT	29-AUG-2002 (first entry)					
XX						
DE	Listeria innocua contig DNA sequence #10.					
XX						
KW	Antibacterial; Listeria; food contamination; mutational analysis;					
KM	infection; ds.					
XX						
OS	Listeria innocua.					
PN	WC200228891-A2.					
XX						
PD	11-APR-2002.					
XX						
PF	04-OCT-2001; 2001MO-FR03061.					
XX						
PR	04-OCT-2000; 2000FR-0012697.					
XX						
PA	(INSP) INST PASTEUR.					
PA	(CNRS) CNRS CENT NAT RECH SCI.					
XX						
PI	Kunst F, Glaser P;					
DR	WPI; 2002-332479/37.					
XX						
PT	New genomic sequences from Listeria species, useful for detection,					
PT	treatment and prevention of infection, also related polypeptides,					
XX	antibodies and modulators .					
PS	Claim 5; SEQ ID 10; 180pp: French.					
XX						
CC	The present invention relates to nucleic acid sequences					
CC	(ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes					
CC	and primers for identification and/or detection of Listeria (e.g. as					
CC	contaminants in foods, or mutational analysis) and for analysis of					
CC	gene expression. Proteins encoded by the nucleic acid sequences can be					
CC	used to screen for compounds that modulate gene expression, replication					
CC	and pathogenicity of Listeria (potential therapeutic agents), also for					
CC	treating infections by Listeria, and are useful as immunogens in					
CC	anti-Listeria vaccines.					
CC	Note: The sequence data for this patent did not form part					
CC	of the printed specification, but was obtained in electronic format					
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.					
XX						
SQ	Sequence 1163020 BP; 388339 A; 197589 C; 235612 G; 341474 T; 6 other:					
Query Match						
Best Local Similarity 49.8%; Pred. No. 43;						
Matches 130; Conservative 0; Mismatches 125; Indels 6; Gaps						
OY	151	ATTGCTGCTATTGGAACAATCCGAAACAAATGATCCATGAAAGTACCTGATCGCGAT	210			
Db	944235	ATTAGTCACATGAAAGATCGGAAAGAAAGCAATTAATTTATGCGTAGGTTGTGACGTT	944176			
OY	211	AAATGCAAGCTCATGTTGGTATTTCGATTTCAAGAGTCAATTAGCCATGCAAAACATT	270			
Db	944175	GCTGCAAAAGAAATTAATAATTTGTTGACCCCTCTGTGCTTTAGCGCGCAAAACATT	944116			

Oy	271	GAGCTGCAGGATTTAAACAACAAATGAACGTCAAGTGTATGTTCAAAGTGGAAGAG	330
Dc	944115	GAGCG-----AACAAATATGATTATCCGGCTAGTTGGTAAGCAGTGAAGCGGAGCAG	944062Z
Oy	331	GGTATTGTTAAAGTCATATTTGGTAGTCGGTTCACGATGATATCGTTCGATGAATAT	390
Dc	944061	TATTTCTGTTCAACGTGACTTACTCAAAGATATTTGCCAAGCAGCTCGCAAAAACGAATTT	944002Z
Oy	391	GATTTAGCATACAAATTTGGGT 411	
Dc	944001	GAATTACCACCTTAATTTTGT 943981	
<hr/>			
RESULT 32			
xx	AAQ40697	ID	AAQ40697 standard; DNA; 1005 BP.
xx	AAQ40697;		
xx	23-AUG-1993	(first entry)	
Dc	B. burgdorferi strain Ip90 OspB coding region.		
KW	OspB: coding region; B. burgdorferi; strain; B31; North America;		
KM	tick; vector; Ikodes damini; primer; ACAI; Ip90; class; I; II;		
KW	III; detection; ss.		
xx	Borrelia burgdorferi.		
xx	Key	Location/Qualifiers	
FH	CDS	6..896	
FT	primer_bind	/tag= a	
FT		1..9	
FT		/tag= b	
FT	primer_bind	/note= "Detection primer, claim 5"	
FT		762..773	
FT		/tag= c	
FT		/note= "Detection primer, claim 5"	
xx	MO308306-A.		
xx	29-APR-1993.		
xx	22-OCT-1992;	92MO-US08972.	
xx	22-OCT-1991;	91US-0779185.	
xx	(SYMB-) SYMBICOM AB.		
PA	Barbour AG, Bergstrom S, Hansson L;		
PI	WPI; 1993-152497/18.		
DR	P-PSDB; AAR35439.		
PT	Detecting Borrelia burgdorferi by PCR-DNA analysis - using		
PT	defined nucleotide primer sequences, also a vaccine against Lyme		
PT	disease		
xx	Disclosure; Fig 3b; 11pp; English.		
xx	This sequence represents the OspB coding region of the B. burgdorferi		
xx	strain Ip90. The strain Ip90 is a Soviet union isolate from the tick		
xx	vector Ixodes persulcatus. The primer sequences indicated can be used		
xx	In a method to detect the presence of B. burgdorferi in an animal.		
xx	These primers correspond to regions within the B. burgdorferi genome		
xx	which are identical between strains B31, ACAI and Ip90 (see also		
xx	AHQ40696-967). These strains of B. burgdorferi are representatives of		
xx	the three classes of B. burgdorferi species, class I, II and III. The		
xx	detection method is very sensitive and can be used to detect B.		
xx	burgdorferi infection at a very early stage, even before clinical		
xx	signs are manifest.		

KW skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
OS Staphylococcus aureus.
XX EP786519-42.
XX 30-JUL-1997.
XX 07-JAN-1997; 97EP-0100117.
XX 05-JAN-1996; 96US-0009861.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX WPI: 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
XX Claim 1: Page 822; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
SQ Sequence 1021 BP; 386 A; 135 C; 201 G; 295 T; 4 other:
Query Match 5.0%; Score 38.4; DB 18; Length 1021;
Best Local Similarity 50.0%; Pred. No. 8.5;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 12 AATTAACATCCAAACAAATATCTTACAAATGATGAATTTTGTGATTCGCGCGT 71
DB 710 AATGCTATTCAAAATTCATTTTAATTAATGATTAATATGCGTGAAGCATGTGTGT 769
QY 72 GGCATTTCGCGGTTTCGCGTATCCAAATTCATGATGAATTAATCCGAAGAATCA 131
DB 770 TGATGATTAAGATGTTTACTCTGATCATATACTATGTTCAATCTTAACAGCTCTGA 829
QY 132 CAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191
DB 830 AAAAGATTAGGCGTGTGCTATGATGATGATGATGATGATGATGATGATGATGAT 889
QY 192 AGTACTGATCA 203
DB 890 TGAACGCGGTGA 901
RESULT 37
ID ABN91558 standard; DNA; 1170 BP.
XX
XX ABN91558;

XX 24-JUL-2002 (first entry)
DT Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1021.
XX
DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX Staphylococcus epidermidis; gene therapy; gene; ds.
KW antibacterial; gene therapy; gene; ds.
XX
OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
PI
XX WPI: 2002-381255/41.
XX P-PSDB: ABP39013.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections
XX
XX Disclosure: SEQ ID 1021; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 1170 BP; 407 A; 160 C; 241 G; 362 T; 0 other:
Query Match 5.0%; Score 38.4; DB 24; Length 1170;
Best Local Similarity 50.5%; Pred. No. 8.9;
Matches 93; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 13 ATTAACATCCAAACAAACATATCTTACAAATGATGAATTTTGTGATTCGCGCGT 72
DB 493 ATGCGCATCCAAACAAATCTTTTAATCATATGATTAATATGTTGAGCGATGTGT 552
QY 73 GCATTTCGCGGTTTCGCGTATCCAAATTCATGATGAATTAATCCGAAGAATCAAC 132
DB 553 GAGCTTTAGATGTTTATCTATGATGATGATGATGATGATGATGATGATGATGAT 612
QY 133 AAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
DB 613 AAAAATTTAGGCGCTTGTGTTATGATGATGATGATGATGATGATGATGATGATGAT 672
QY 193 GTAC 196
DB 673 GAAC 676
RESULT 38
ID AAS52152 standard; DNA; 1194 BP.
XX
XX AAS52152;
XX
XX 13-FEB-2002 (first entry)

XX Staphylococcus aureus DNA for cellular proliferation protein #569.
DE
XX
XX Antisense: ds: prokaryotic cellular proliferation gene;
KM antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
OS
XX WO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PE
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAR-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR P-PSDB: AAU34293.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Claim 27; Seq ID No 4734; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1194 BP; 433 A; 152 C; 245 G; 364 T; 0 other;

Query Match 5.0%; Score 38.4; DB 23; Length 1194;
Best Local Similarity 50.0%; Pred. No. 8.9;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 12 AATAACAATCCAAAAAACAATATCTTACAAATGATGATGATTTTGGTGGCGGT 71
DB 471 AATTGCTATTCAAAAATGATTTTAAATTAATGATTAATCCGTGAAGCATGTGGTGT 530

QY 72 GGCATTTGCGCGCTGGCTGATCATCAATTCATGATTAATAATCCAGGAATCA 131
DB 531 TGATGTATATAGATGTTTACTGTGATGACATATAAGTATGTTCAACGACTACTGA 590

QY 132 CAACCTATTGATGATGCCATTGCTGCTATTGTAACAATCCGAACAAATATCAATGAA 191
DB 591 AAAAGATTAGTGCATGTGTCATTTGATTTGTTGTAAGACGTTAAGCAAGTCTTTT 650

QY 192 AGTACCTGATCA 203
DB 651 TGACGCCGTGA 662

RESULT 39
AAD00096/C
ID AAD00096 standard; DNA: 1401 BP.
XX
XX AAD00096;
AC
XX
XX 31-JUL-2000 (first entry)
DT
XX
XX
DE Human zsig49 degenerate DNA.
XX
XX Human zsig49 cDNA: human chromosome 1q24; metabolic disease; diabetes;
KM pancreatic disease; reproductive development; testicular function;
KM wound healing; anti-microbial; antidiabetic; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX WO200023591-A2.
PN
XX
XX 27-APR-2000.
PD
XX
XX 20-OCT-1999; 99WO-US24579.
PE
XX
XX 21-OCT-1998; 98US-0176545.
PR
XX
XX (ZYMO) ZYMOGENETICS INC.
PA
XX
XX Sheppard PO, Holly RD, Gao Z, Whitmore TE, Maurer MF;
PI
XX
XX WPI: 2000-339690/29.
DR
XX
XX zsig49 polypeptides and polynucleotides, useful for treatment and
PT diagnosis of metabolic diseases such as diabetes -
XX
XX
PS Disclosure: Page 125; 133pp; English.
XX
XX The present sequence is a degenerate DNA sequence that encompasses all
CC DNAs that encode the human zsig49 protein. zsig49 is expressed in
CC organs of the endocrine system, pancreas, testis, thymus, adrenal gland,
CC thyroid, and pituitary gland. This gene was localised to human chromosome
CC 1q24 which is also the location of a susceptibility locus for prostate
CC cancer (HPC1). zsig49 genes and proteins are useful for treatment and
CC diagnosis of metabolic diseases such as diabetes, pancreatic conditions.
CC They may exert regulatory effects on male gametes, reproductive
CC development and testicular functions. They are therapeutically useful for
CC aiding digestion, wound healing and anti-microbial applications.
XX
SQ Sequence 1401 BP; 275 A; 112 C; 229 G; 188 T; 597 other;

Query Match 5.0%; Score 38.4; DB 21; Length 1401;
Best Local Similarity 27.2%; Pred. No. 9.3;
Matches 82; Conservative 57; Mismatches 162; Indels 0; Gaps 0;

QY 461 TTGCCTTGCCCTTGAATTTCTGATGAAGTAAACATATGACATCTTTGAAGTAC 520
DB 443 TTNSWYTRTNSMNAWNGTNSMRTTDTATNGCCNARTRTTTNGCTGYTCRCARTCY 384

QY 521 GACATTTGGCTTAATGTTGCAACCATATTGGTGGCTTTCATCTGATCCAAATTTTG 580
DB 383 TGNARTTCTTTTNGTTRTTTNGCCATTTTNNARRATTTTTRTNSWTRGNCNCYTTDATN 324

QY 581 GCGTTTATCTGATGATTAAGACCGCTATTTCACAGACACCGTACGTAAAGAAATGACCA 640
DB 323 CCNACNGDATATTTTNSWTRCNGRANGNGYTCNACRTGRCARTCTTTATNSWRCARTGN 264

QY 641 AAGTATTTGGACACCATTTTAAGCGTAATTTGAAAAAATTTAAACATATGACATATTT 700
DB 263 CCRTCYTNGGCCACARAANGRCANGCNKYTTNARRTTNACYTTRTARTANCKRAARTAR 204

QY 701 TTCCAACTGACATCTCTATTTCACGACATPAATAAATTTTATTATTTCGC 760
 Db 203 TCNCKTCTCGNARATTTTADATTTNGRAADATYTRFRANGRTTRARTRTCDATN 144
 QY 761 C 761
 Db 143 S 143

RESULT 40
 AAA9396
 ID AAA9396 standard; DNA; 1413 BP.
 XX
 AC AAA9396;
 DT 15-SEP-2000 (first entry)
 XX
 DE Staphylococcus aureus monomeric FtsA nucleotide sequence SEQ ID NO:3.
 XX
 KM Staphylococcus aureus; monomeric FtsZ; monomeric FtsA; antibacterial;
 KM Escherichia coli; multimeric protein; bacterial infection; vaccine;
 KM microbial infection; screening; ds.
 XX
 OS Staphylococcus aureus.
 XX
 PN WO200024252-A1.
 XX
 PD 04-MAY-2000.
 XX
 PF 20-OCT-1999; 99WO-US24653.
 XX
 PR 23-OCT-1998; 98US-0105315.
 PR 19-APR-1999; 99US-0129965.
 PR 03-MAY-1999; 99US-0132333.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 XX
 PI Yan K, Pearce KH;
 XX
 DR WPI: 2000-350517/30.
 DR P-PSDB; AAY91109.
 XX
 PT Multimeric FtsA polypeptide, useful as a vaccine for treating
 PT bacterial infections and for diagnosing diseases associated with
 PT bacterial infection
 XX
 PS Example 1; Page 3-4; 45pp; English.

CC The present invention describes isolated multimeric FtsZ:FtsA
 CC polypeptides (1) comprising the protein sequences given in AAY91108,
 CC AAY91109, AAY91110 and AAY91111. Also described are: (1) an antibody
 CC (Ab) specific for (1); (2) a method (M1) for treating an individual
 CC requiring enhanced or reduced activity or expression of (1) comprising
 CC administering an agonist or antagonist to (1), respectively; (3) a
 CC method (M2) for diagnosing a disease or susceptibility to a disease
 CC related to expression or activity of (1) comprising determining the
 CC presence or absence of a mutation in the nucleotide sequence encoding
 CC (1) and/or analysing for the presence or quantity of (1) in a sample;
 CC and (4) a method (M3) for screening to identify compounds that activate
 CC or inhibit the function of (1) selected from: (a) measuring the binding
 CC of a candidate compound to the polypeptide using a label; (b) measuring
 CC the binding of a candidate compound to (1) in the presence of a labeled
 CC competitor; (c) testing whether the candidate compound results in a
 CC signal generated by activation or inhibition of (1); or (d) detecting
 CC the effect of a candidate compound on the production of mRNA encoding
 CC (1) using e.g. enzyme linked immunosorbent assay (ELISA). AAY91108 to
 CC AAY91111 are useful as vaccines for treating microbial infections such
 CC as Staphylococcus, Streptococcus and Escherichia coli infections. The
 CC polypeptides are also useful for screening for antibacterial compounds
 CC and for detecting diseases associated with microbial infections. The
 CC present sequence encodes Staphylococcus monomeric FtsA.
 CC
 XX Sequence 1413 BP; 546 A; 175 C; 278 G; 414 T; 0 other;

Query Match 5.0%; Score 38.4; DB 21; Length 1413;
 Best Local Similarity 50.0%; Pred. No. 9.3;
 Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 12 AATACAAATCCAAAAACATATCTTACAAATGATGAATTTTGTATGTCGCCGT 71
 Db 471 AATTCGTAATTCAAAAATCCGATTTTAAATGATTAATGCGTGAAGCATGTCGT 530
 QY 72 GGCAATTTGCGCGCTTCGCGTGAATTCACATATGATTAATGACGGAATTCAA 131
 Db 531 TGATGTAATTAAGATGTTACTGTGATCAATTAATGTTCAATCCTAACAGCTACTGA 590
 QY 132 CAAGCTATTTGATGATGCCATTTGCTCTATTGAAACATCCGAACATAGATCAATGAA 191
 Db 591 AAGAAGTTAGTGTCATGTCATGTCATTAATGTCGTAAGACGTTACCGCAAGTTGCTTTTA 650
 QY 192 AGTACCTGATCA 203
 Db 651 TGAACGCGGTGA 662

RESULT 41
 AAA91088
 ID AAA91088 standard; DNA; 1413 BP.
 XX
 AC AAA91088;
 XX
 DT 20-APR-2001 (first entry)
 XX
 DE S. aureus FtsA coding sequence.
 XX
 KM FtsA; cell division protein; otitis media; conjunctivitis; pneumonia;
 KM bacteraemia; meningitis; sinusitis; pleural empysema; endocarditis;
 KM Helicobacter pylori; infection; cancer; gastrointestinal carcinoma;
 KM gastric ulcer; gastritis; therapy; ds.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1413
 FT /tag= a
 FT /product= "FtsA"
 XX
 PN WO200100649-A1.
 XX
 PD 04-JUN-2001.
 XX
 PF 27-JUN-2000; 2000WO-US17566.
 XX
 PR 29-JUN-1999; 99US-0342459.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Yan K, Pearce K;
 XX
 DR WPI: 2001-122994/13.
 DR P-PSDB; AAY97598.
 XX
 PT Isolated cell division family (FtsA) polypeptide of Staphylococcus
 PT aureus, useful for treating bacterial infections such as otitis media,
 PT pneumonia, conjunctivitis, bacteraemia, meningitis, sinusitis,
 PT endocarditis -
 XX
 PS Claim 2; Page 3; 38pp; English.
 XX
 CC This sequence encodes the Staphylococcus aureus FtsA protein of the
 CC invention. FtsA is a protein from the cell division family. FtsA is
 CC useful to treat an individual in need of enhanced activity or expression
 CC of immunological response to FtsA by administering an agonist to FtsA,
 CC or to treat an individual having need to inhibit activity or expression
 CC of FtsA by administering an antagonist to FtsA or a nucleic acid molecule

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147382.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 5.0%; Score 38.2; DB 21; Length 630;
Pred. No. 8.4;
Matches 136; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 60 GATTCGCGCCGTCGATTTGCGCCGTTGGCTGCATGCATTTACTATGATAAATCAC 119
DB 206 GCTCGTTCCTCCACCCTATGATCTACACACACCCATTCACAAAGGAAT 265
QY 120 CGAAGAAATCAACAAGCTATGATGATGCCATGCTGATTAACAATCCGAACAT 179
DB 266 AGTACTCTGAAATTAAGCAATTTGAAGCATGATCTACTGCAATTTAGAACAAAGGTTAA 325
QY 180 AGATCAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGATTTGGA 239
DB 326 ACTGTGCTTAAAGAGAGTGCTCGCGGAGTGCAGCGTGCATGACAAATTTGCTGCTGA 385
QY 240 TTTCAAGGTCGATTTACCATGCGAAACATTTGAGGCTCGAGGATTTGAACAATGAACG 299
DB 386 GCACATGCTTAAGCTTAAGATGATGAATGAAGATGATGATGAGGAGGATTAATTTG 445
QY 300 TCAAGGTGATGCTAATGCAAGGTGAAGAGGATTTGTTAAAGCTATTTGTTGATCG 358
DB 446 TGAAGAGAAACACACTGCAATGGAGAGATGATGATGAGAGTACCGGATTAATTTG 504

RESULT 46

AAV74654 standard; DNA; 3150 BP.

AAV74654:

16-MAR-1999 (first entry)

Staphylococcus aureus contig SEQ ID #343.

Computer readable medium; vaccine; S.aureus infection; immunodetection;
cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
skin infection; surgical wound infection; scalded skin syndrome;
toxic shock syndrome; ds.

Staphylococcus aureus.

Location/Qualifiers
Key 481..540
misc_feature

/*tag- a
/note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

misc_feature 2281..2340
/*tag- b
/note=

"these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

EP786519-A2.

30-JUL-1997.

07-JAN-1997; 97EP-0100117.

05-JAN-1996; 96US-0009861.

(HUMA-) HUMAN GENOME SCI INC.

Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

PI Rosen CA:
XX WPI: 1997-374922/35.
DR
XX
XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
XX
PS Claim 1: Page 1220-1222; 3271pp; English.
XX
XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
SQ Sequence 3150 BP; 983 A; 390 C; 608 G; 1049 T; 120 other;
Query Match 5.0%; Score 38.2; DB 18; Length 3150;
Best Local Similarity 49.7%; Pred. No. 13;
Matches 97: Conservative 0; Mismatches 98; Indels 0; Gaps 0;
OY 562 ATCTTGATCAATTTTGGCTTTATGTGATGATGACCGCTATTTTCCAGACACC 621
DB 2597 ATACTGTGACAAATATTAAGTGTACAGCAGTGTAAATTTTATGATGACGCTTT 2656
OY 622 GTACGTAAGGAATGACCAAGATTTGGACACGACATTTAAACGTGATGAAATAAT 681
DB 2657 ATTATTTGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2716
OY 682 TAACCAATAGACATCTTTTCCAACTGTACATCTCTATTTCACTGACATAAATAAA 741
DB 2717 ATAGATTTAGACAAGATTTTACGCGTGAACATGCTTACTTTGAAGGTAGCTAATAA 2776
OY 742 ATTTTATTTTATTT 756
DB 2777 TTCAATTAACATATTT 2791
RESULT 47
AAV11344
ID AAV11344 standard; DNA; 4200 BP.
XX
XX AAV11344:
AC
XX
XX 28-AUG-1998 (first entry)
DE
XX
XX Group B streptococcal C-beta protein DNA.
KW
XX
XX C-beta protein; beta antigen; vaccine; group B Streptococcus; ss.
OS
XX
XX Streptococcus sp. strain A909.
FH
XX
XX Key Location/Qualifiers
FT CDS 320..3814
FT sig_peptide 320..430
FT mat_peptide 431..3811
FT /*tag= b
FT /*tag= c

XX
XX WO9809648-A1.
XX
XX 12-MAR-1998.
XX
XX 05-SEP-1997; 97WO-US15319.
XX
XX 06-SEP-1996; 96US-0024707.
XX
XX (NABA-) NORTH AMERICAN VACCINE INC.
XX
XX
XX Blake MS, Tal JT;
PI
XX
XX WPI: 1998-193324/17.
DR
XX
XX P-PSDB: AAM40537.
PT
XX
XX Group B streptococcal C-beta proteins - having amino acid
PT substitutions to reduce immunoglobulin A binding while retaining
PT antigenicity, for use in vaccines
XX
XX
PS Claim 1; Fig 1; 59pp; English.
XX
XX This polynucleotide codes for a group B Streptococcus C-beta
CC protein (see AAM40537), or beta antigen, that binds to human IgA.
CC The invention relates to polynucleotides (see AAV11345-47) encoding
CC mutant C-beta proteins of formula A(X1-X12)B, where A comprises
CC amino acids 1-164 of mature C-beta protein, B comprises amino acids
CC 177 to 1096-1127 of mature C-beta protein, and X1-X12 are each
CC selected independently from Ala, Val, Leu, Ile, Pro, Met, Phe, Trp,
CC a bond, or the corresponding amino acid residue from B-beta protein,
CC provided that at least one of X1 to X12 is other than the wild-type
CC amino acid and that the LPYTG motif may be missing. Vectors
CC comprising such polynucleotides, and transformed host cells, are
CC claimed. The mutant C-beta proteins have reduced IgA binding while
CC retaining most of the antigenicity of the wild-type protein. They
CC can be used alone, or conjugated to a polysaccharide, in vaccines
CC for use against group B Streptococci (claimed).
XX
SQ Sequence 4200 BP; 1632 A; 685 C; 818 G; 1065 T; 0 other;
Query Match 5.0%; Score 38.2; DB 19; Length 4200;
Best Local Similarity 52.9%; Pred. No. 14;
Matches 82: Conservative 0; Mismatches 73; Indels 0; Gaps 0;
OY 85 GTTGGCGATGATCAATTTCTATGATTAATAATCACCAGGAATCAACAAGCTATTTGAT 144
DB 680 GATGTGGATTAACCAATTTCTATCTCATGACGAGAAAAATGAGATTAAACAAATTTGAT 739
OY 145 GATGCCATTGCTGTATTTGAAACAATCCGAACATAGATCATGAAGAACTGATCAT 204
DB 740 GAACAAATGATTTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 799
OY 205 GCCGATTAATTTGCAACGTCATGTTGTAATTGGGA 239
DB 800 CTGTTACACATCAACACATGAGAGAGTTGAGAA 834
RESULT 48
ABL32289
ID ABL32289 standard; DNA; 5875 BP.
XX
XX ABL32289:
AC
XX
XX 26-MAR-2002 (first entry)
DE
XX
XX Human immune system associated gene SEQ ID NO: 262.
XX
XX Human: immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antiparasitic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 262; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5875 BP; 1836 A; 27 C; 1078 G; 2934 T; 0 other;
XX
Query Match 5.0%; Score 38.2; DB 24; Length 5875;
Best Local Similarity 56.9%; Pred. No. 16;
Matches 70; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
XX
QY 637 ACCAAGTATTGGCACCAGCATTTAAACGTAATGGAAAAAATTAAACCATGACATC 696
DB 4084 ACGTTAAGTTGGGTAAATAGAGTGAATTCGTTTAAATTAATTAATTAATTAATTA 4143
QY 697 ATTTTCACTGTACATCTCTATTCACAGCATTAATAATTAATTTATTTATTTATTT 756
DB 4144 TTTTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATTT 4203
QY 757 TCT 759
DB 4204 TTT 4206
XX
RESULT 49
AAV52498/c
ID AAV52498 standard; DNA; 1436 BP.
XX
XX AAV52498;
XX
XX 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:365.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
XX Streptococcus pneumoniae.
XX
XX WO9818931-A2.
XX
PD 07-MAY-1998.
XX

PF 30-OCT-1997; 97WO-US19588.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
XX
PS Claim 1; Page 1380-1381; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 1436 BP; 357 A; 465 C; 331 G; 282 T; 1 other;
XX
Query Match 5.0%; Score 38; DB 19; Length 1436;
Best Local Similarity 43.0%; Pred. No. 12;
Matches 185; Conservative 0; Mismatches 245; Indels 0; Gaps 0;
XX
QY 63 TCGTCCGCTGGCATTTTGCCCGTTTGCGCTGATCCATTCATGATGATTAATTCACCGA 122
DB 765 TGAAGCGGACGACCTCGTGTATGCTGAGGCGCTAGTACTTGTGAGGCTGACGC 706
QY 123 AGAAATCAACAAGCTATTGATGATGCCATTGGCATTTGAACAATCCGAACAATAGA 182
DB 705 ACTGTTGATGCCGATTCAGATGACTGTGTAAGGCCGACGACCTGTCCACGCTGAGGC 646
QY 183 TCCAAATGAAGTACCTGATCATGCGGATTAATTCGAACGTCATGTGTTGATTTGATTT 242
DB 645 TGAAGCACTGCTACTTCTGCTGAGCGCACTGTCATCTTGTGCAACCTGAGGCGCTGT 586
QY 243 CAAAGTGCAATTAGCCATGCGAAACATTGAGGCTCGAGGATTGAACAATGAAGCTCA 302
DB 585 CGATGCTGAGGCTGACCACTGTTGATGCAAGTTCGAAGCTGATGCTGCTGAGGC 526
QY 303 AGGTGATGCTAATGTCAAAGGTGAAGGGGATTTGAAGCATTCATTTGTTGATCGGTGT 362
DB 525 TGACGTACTGTTGATGCTGATCAGACGACATCGTTGAGCGCTGAAGCCAGACACTTGT 466
QY 363 TCACGATGATATGCTCGATGGAATGATGATTAAGCATTAACAATTTGGGTGATCTTCA 422
DB 465 TGACGCTGATGCTGAGGCGCTAGTACTTGTGTAAGGAGGAGCGACTGTTGATGCGCAT 406
QY 423 AACCATCATGTCATTTGGATTAAGATTTTGTGTTGTTGCTTGTCCCTTGAATTTTC 482

Db 405 AGACCGCTGAGACTTCTGCGAAGCTGAGCACTTGTGATGCTGAAGCTGATGCACTCGT 346
 Oy 483 TGATGAGGT 492
 ||||| |
 Db 345 TGATGCTGAT 336

RESULT 50

ABN92642
 ID ABN92642 standard; DNA: 1722 BP.

ABN92642;

24-JUL-2002 (first entry)

Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2105.

Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 antibacterial; gene therapy; gene; ds.

Staphylococcus epidermidis.

US6380370-B1.

30-APR-2002.

13-AUG-1998; 98US-0134001.

14-AUG-1997; 97US-055779P.

08-NOV-1997; 97US-064964P.

(GENO-) GENOME THERAPEUTICS CORP.

Doucette-Stamm LA, Bush D;

WPI: 2002-381255/41.

P-PSDB: ABP40097.

Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 polypeptide, useful for diagnosing and treating bacterial infections -

Disclosure; SEQ ID 2105; 267pp; English.

ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 frame (ORF) nucleic acid sequences which encode the amino acid sequences
 given in ABP5124 to ABP37960. The S. epidermidis sequences have
 antibacterial activity and can be used in gene therapy. The sequences
 can also be used in the diagnosis and treatment of bacterial infections,
 particularly S. epidermidis infections. The sequences can be used to
 screen for compounds able to interfere with the S. epidermidis life
 cycle or inhibit S. epidermidis infection.
 N.B. The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from the
 USPTO web site.

Sequence 1722 BP; 620 A; 235 C; 355 G; 512 T; 0 other.

Query Match 5.08; Score 38; DB 24; Length 1722;

Best Local Similarity 49.5%; Pred. No. 12;
 Matches 98; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Oy 263 GAACATTTGAGCGTCGAGATTGAACAATGAACGTCAGTCGATCTAATGTCAAG 322
 ||||| |
 Db 41 GTAACTTTTGTAGTTTAAATAAATAAATAAAGATATTGTATTCACATCGGTG 100
 Oy 323 GTGAAGAGGTATTGTTAAAGCTCATTTGTTGATCGGTTCACGATGATGCTCGA 382
 ||||| |
 Db 101 GAGTAGGCGGAATCGCTAAATAATATGTACATCGTTGAAGTAGACGATGAATGTTATGT 160
 Oy 383 TGAATATGATTTAGCATACAAATTTGGGTATCTTCATCCAAACCATCATGTCATTTCGG 442
 ||||| |
 Db 161 TGGATGCGAGATTATATGTTCCAGAAAGATGAATGCTTGGGTAGATATCGTTATTCTCTG 220

Oy 443 ATATTCAAGATTGTTGTG 460
 ||||| |
 Db 221 ACATTCAATATGTCAATTG 238

Search completed: February 20, 2003, 20:55:59
 Job time : 2559 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 20:13:26 : Search time 80 Seconds
(without alignments)
2917.263 Million cell updates/sec

Title: US-10-024-955-6

Perfect score: 761

Sequence: 1 GATCTATATCATACATACAAAT.....ATTATTATTTTATTTTCGCC 761

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/1na/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/1na/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/1na/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/1na/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/1na/PCUTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/1na/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	761	100.0	761	4	US-08-553-336A-6	Sequence 6, Appli	
	2	518.6	68.1	812	4	US-08-553-336A-1	Sequence 1, Appli	
	3	517	67.9	812	3	US-08-462-778-1	Sequence 1, Appli	
	4	47.8	6.3	1386	4	US-09-134-001C-2474	Sequence 2474, Appli	
	5	40	5.3	2538	1	US-07-839-433-1	Sequence 1, Appli	
C	6	39.6	5.2	1702	4	US-08-961-527-367	Sequence 367, Appli	
C	7	39.6	5.2	32768	4	US-08-961-527-71	Sequence 71, Appli	
C	8	39.4	5.2	276	4	US-09-134-001C-527	Sequence 527, Appli	
C	9	39	5.1	1350	4	US-09-134-001C-2836	Sequence 2836, Appli	
	10	38.8	5.1	1958	1	US-08-137-175A-7	Sequence 7, Appli	
	11	38.8	5.1	1958	3	US-08-479-017-7	Sequence 7, Appli	
	12	38.4	5.0	1170	4	US-09-134-001C-1021	Sequence 1021, Appli	
	13	38.4	5.0	1413	4	US-09-342-459-1	Sequence 1, Appli	
C	14	38	5.0	1436	4	US-08-961-527-365	Sequence 365, Appli	
C	15	38	5.0	1722	4	US-09-134-001C-2105	Sequence 2105, Appli	
	16	37.2	4.9	1776	4	US-09-134-001C-438	Sequence 438, Appli	
	17	37	4.9	7832	4	US-09-004-838-94	Sequence 94, Appli	
C	18	36.6	4.8	932	1	US-08-242-932-7	Sequence 7, Appli	
	19	36.6	4.8	932	1	US-08-714-481-7	Sequence 7, Appli	
	20	36.6	4.8	932	5	PCT-US95-06111-7	Sequence 7, Appli	
	21	36.6	4.8	2617	1	US-08-430-024-1	Sequence 1, Appli	
	22	36.6	4.8	2617	1	US-08-782-009-1	Sequence 1, Appli	
	23	36.6	4.8	2617	3	US-09-017-302-1	Sequence 1, Appli	
	24	36.6	4.8	3294	4	US-08-923-992A-7	Sequence 7, Appli	
	25	36.6	4.8	3312	4	US-08-923-992A-3	Sequence 3, Appli	
	26	36.6	4.8	3384	4	US-08-923-992A-5	Sequence 5, Appli	
	27	36.6	4.8	3730	1	US-08-242-932-8	Sequence 8, Appli	

28	36.6	4.8	3730	1	US-08-714-481-8	Sequence 8, Appli
29	36.6	4.8	3730	5	PCT-US95-06111-8	Sequence 8, Appli
30	36.6	4.8	4200	1	US-08-242-932-1	Sequence 1, Appli
31	36.6	4.8	4200	1	US-08-714-481-1	Sequence 1, Appli
32	36.6	4.8	4200	4	US-08-923-992A-1	Sequence 1, Appli
33	36.6	4.8	4200	5	PCT-US95-06111-1	Sequence 1, Appli
34	36.6	4.8	5495	1	US-08-602-010A-1	Sequence 1, Appli
35	36.6	4.8	5495	1	US-08-602-010A-2	Sequence 2, Appli
36	36.6	4.8	5495	1	US-08-680-726A-1	Sequence 1, Appli
37	36.6	4.8	5495	1	US-08-680-726A-2	Sequence 2, Appli
38	36.6	4.8	5495	3	US-09-092-409-1	Sequence 1, Appli
39	36.6	4.8	5495	3	US-09-092-409-2	Sequence 2, Appli
40	36.6	4.8	10592	1	US-08-680-726A-51	Sequence 51, Appli
41	36.6	4.8	10592	1	US-08-680-726A-52	Sequence 52, Appli
42	36.6	4.8	10592	3	US-09-092-409-51	Sequence 51, Appli
43	36.6	4.8	10592	3	US-09-092-409-52	Sequence 52, Appli
44	36.6	4.8	11091	4	US-09-134-001C-2243	Sequence 2243, Ap
45	36.4	4.8	1068	4	US-08-961-527-373	Sequence 373, App

ALIGNMENTS

RESULT 1
US-08-553-336A-6
Sequence 6, Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
NUMBER OF INVENTIONS: House Dust Mite and Uses Therefor
CORRESPONDENCE ADDRESSES: 15
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 761 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 43...681.
US-08-553-336A-6
Query Match 100.0%; Score 761; DB 4; Length 761;
Best Local Similarity 100.0%; Pred. No. 4.6e-193;
Matches 761; Conservative 0; Mismatches 0; Indels 0;

Db 568 TCATATGCTGGCTCTTCAATTTTGGATCCAAATTTGCTGCTTAVCCGATGTTTGAC 627
 QY 603 CCCTATTTTCCAAAGACCCGTAAGTAAGAAATGACCAAGTATTGGACGACATTTTAA 662
 Db 628 CCGTATTTTCCAGGATACCGTACGTCAGCAAAATGACCAAGATTTGGACGACGATTTCAA 687
 QY 663 AGCTGATTTGGAAAAATTAACCAATAGACATCTTTTCCAACTGTACAACTCTAT 722
 Db 688 AAAAGAAATTTGGAACGAAA-CAACCAATAGACTTACACACACATTAACACTGTTATTTTAA 746
 QY 723 TCACGTGACATTAATAATTTTATTTTAT 755
 Db 747 CACTGATATCAAAATGAATTTTAT 779

RESULT 3

US-08-462-778-1
 ; Sequence 1, Application US/08462778
 ; Patent No. 6077517
 ; GENERAL INFORMATION:
 ; APPLICANT: Thomas, Wayne R.
 ; TITLE OF INVENTION: Allergenic Protein and Peptides From
 ; TITLE OF INVENTION: House Dust Mite and Uses Therefor
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lablue & Cockfield
 ; STREET: 60 State Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/462,778
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: USSN 08/031,141
 ; FILING DATE: 12 March 1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Amy E. Mandragouras
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-227-7400
 ; TELEFAX: 617-227-5941
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 812 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 68..712
 ; US-08-462-778-1

Query Match 67.9%; Score 517; DB 3; Length 812;
 Best Local Similarity 81.3%; Pred. No. 2.8e-128;

Matches 612; Conservative 0; Mismatches 140; Indels 1; Gaps 1;

QY 3 TCCTATATCAATAACCAAAATATCTTACAAATGATGAATTTTGTGTAT 62
 Db 28 TTTTTCATATCGTAAATATCAATTCACCTTTTACCAAAATGATGAATTTATATGAT 87
 QY 63 TCGTCGCGTGGCATTTGTCGCGTTTCGCTGATCCAAATTCATATGATAAATACCGA 122
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 88 TCGTCGCGGACGCTTTTGTGTCGCTTGGCTGATCCAAATTCATATGATAAATACCGA 147
 QY 123 AGAAATCAACAAGCATATGATGATGCATTCGCTGATTTGAACAATCCGAACAATAGA 182
 Db 148 AGAAATTAACAAGCTGTTGATGAAAGCCGTCGCTGAATTTGAAAAATCGAAACATTCGA 207
 QY 183 TCCAAATGAAGTACCTGATGATGCGGATTAATTCGAAGCTCATGTTGATTTGATTT 242
 Db 208 TCCAAATGAAGTACCGGATCATTTGATTAATTCGAACACATATTTGATATCATGATTT 267
 QY 243 CAAAGTCAATTAAGCATGCGCAACATTTGAGGCTGAGAGTTGAACAATCAATCAACGTCA 302
 Db 268 AAAAGTCAATTAAGCATGCGCAACATTCACGTTGAGGATTAACAAATCAATCAACGTGT 327
 QY 303 AGGTGATCTAATGTCAAAGGTAAGAGGCTTGTGTAAGCTCATTTGTTGATGGTGT 362
 Db 328 AGGTGATCTAATGTCAAAGGTAAGAGGCTTGTGTAAGCTCATTTGTTGATGGTGT 387
 QY 363 TCACGATGATATGCTGTCGATGTAATGATTTAGCATCAATTTGGGTGATCTTCAATCC 422
 Db 388 TCATGATGACGTTGTTTCAATGGAATATGATTTAGCATCAATTTGGGTGATCTTCAATCC 447
 QY 423 AACCACTCATGTCATTTGCGAATATTCAGATTTTGTGTCCTTGCCTTGAATTTTC 482
 Db 448 AACCACTCATGTCATTTGCGAATATTCAGATTTTGTGTCGATTAATTCGTCGAAGTTAG 507
 QY 483 TCATGAAGTATACATTAACAATGACATCTTTTGAAGTACACATTCGTAATGTGTCA 542
 Db 508 CGAAGAGGTAATGATGACATTCGTCGAAGTACATTTGCGCAATGTTGCA 567
 QY 543 CCATATTTGCTGCTTTTCATCTTGAATTCATTTTGGCTTATATGATGATTAAGC 602
 Db 568 TCATATGCTGGCTCTTTCATTTTGGATTCATTTTCCGCTGCTTATCGATGTTTGAC 627
 QY 603 CCCTATTTTCCAAAGACCCGTAAGTAAGAAATGACCAAGTATTGGACGACGATTTAA 662
 Db 628 CCGTATTTTCCAGGATACCGTACGTCAGCAAAATGACCAAAATGATTTGGACGACGATTCGA 687
 QY 663 AGCTGATTTGGAAAAATTAACCAATAGACATCTTTTCCAACTGTACAACTCTAT 722
 Db 688 AAAAGAAATTTGGAACGAAA-CAACCAATAGACTTACACACACATTAACACTGTTATTTTAA 746
 QY 723 TCACGTGACATTAATAATTTTATTTTAT 755
 Db 747 CACTGATATCAAAATGAATTTTAT 779

RESULT 4

US-09-134-001C-2474
 ; Sequence 2474, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucelte-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 2474
 ; LENGTH: 1386
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-2474

Query Match 6.3%; Score 47.8; DB 4; Length 1386;
 Best Local Similarity 49.8%; Pred. No. 0.0012;
 Matches 121; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

OY	87	TTGGCGTATCATTCACGTGATTAAATAACCAGGAAGAATCAACAACCTATTGTGA	146
Db	414	TTTTGTAGAACCTCTGTAATTTOTAAAGAAACAAAAGTAGTAAATTTAABACAAAAACAC	473
OY	147	TGCCATTGCTGCTATTTGAAACANTCCGAAACAATAGATCCAATGAAGTACTGATCATGC	206
Db	474	TAAACCTAAACCTGTAAAAAAGTTAAAAAAGAAAAAGTATGCTCTTGCTGATGG	533
OY	207	CGATTAATTCGACGTCACTGTGGTATTGGATTTGCAATTTCAAAGGTCAATATACCGACAA	266
Db	534	GTAATATGATATCCTGCTGCTATTGTTAATGGTTATATACGAACGTCATTTTATMAAAAAA	533
OY	267	CATTGAGGCTCGAGGATTGAAACAATAAGAAACGTCAAGGTGATGCTAATGTCAAAGGTGA	326
Db	594	CATTGTTGATATGATATACCAAAATATTGAAAAGATGACAGCACATCTGTGGTATATATGG	653
OY	327	AGA 329	
Db	654	.AAA 656	

Oy	122	AAGAAATCAACAAAGCATTTGATGATGCCATTCCTCTATTGAAACAATCCGAAACAATAG	181
Oy	1427	ATACTCGAATTATCGTTATTGATATATGCCATTCCTCTAAATTGATACTACTACCAACAG	1486
Oy	182	ATCCAAATGAAGTACCTGATCATGCCGATAAATTGCAACGTCATGTT--GGTATTGTGG	238
Db	1487	AGCGTGATTTAATTGGTATGATATGCGACGAACAATTTGATATCTTTTACACGCTCTTCCG	1546
Oy	239	ATTTCGAAGGATTTTGCCATCGCAACAATTTAGCGCTGAGAGATTGAACAAATGAAC	298
Db	1547	GCTACAAAATTTCCAAAGGACCAAAAGATTATTCGATGCGCTTCATGAAGTTCGTGAC	1606
Oy	299	GTCAGGTGATGCTAATGTCAAAAGTGCAAGGGTATTGTAAAGTCATTTGTTGATCG	358
Db	1607	AACACGATTTGATCTGAAAAATACAGAAAGAGGTGAGATGCAACCCCTCTTGATTTGGTCA	1666
Oy	359	GTGTTCAAGATGATATGCTCTCATGSA	386
Db	1667	GCACCTTCCAGAAATGCTCAAGATGA	1694

```

: RESULT 5
: US-07-839-433-1
: Sequence 1, Application US/07839433
: Patent No. 5420021
: GENERAL INFORMATION:
: APPLICANT: MAROGG, JOHN D
: APPLICANT: TOONEN, MARIA Y
: APPLICANT: VERHUE, WALTER M
: APPLICANT: VERRIPS, CORNELIS T
: TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF
: TITLE OF INVENTION: ALPHA-ACETOLACTIC ACID
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN DARBY & CUSHMAN
: STREET: 1615 L STREET, N.W.
: CITY: WASHINGTON, D.C.
: COUNTRY: U.S.A.
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Tape
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/839,433
: FILING DATE: 19920224
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: KOKULIS, PAUL N
: REGISTRATION NUMBER: 16773
: REFERENCE/DOCKET NUMBER: 94713
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 861-3000
: TELEFAX: (202) 822-0944
: TELEX: 248453 CUSH
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2538 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-07-839-433-1

```

```

RESULT 6
US-08-961-527-367/C
: Sequence 367, Application US/08961527
: Patent No. 6420135
: GENERAL INFORMATION:
: APPLICANT: Charles Kunsch
: TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
: NUMBER OF SEQUENCES: 391
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961,527
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 367:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1702 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
US-08-961-527-367

```

	Query Match	Score 40;	DB 1;	Length 2538;
	Best Local Similarity	47.3%;	Pred. No. 0.17;	
Matches	155; Conservative	0;	Mismatches 170;	Indels 3; Gaps 1
OY	62 TTGCGCCGTGCATTTTTCGCCGGTTTGCGCTGCATCAATTACACTAGTAATAAATCACC	G	C	121
Db	1367 TTAGTGGCTTGTTGGTATGACCCAAATGAATATAGAACCTGTACAACGGATGAGAATTG	A	T	1426

[illegible]

Db 1123 ACTGTTGATGCCAGTTCAGATGACACTGTTGAGCCGACGCACTTGTGACGCTGAGGC 1064
QY 183 TCCATGAAAGTACCTGATCATGCGGATAAATTGGAAGTCATGTTGGATTGTGGATT 242
Db 1063 TGAACTACTGACTGCTGAGGCTGACGCACTGCTACTGCTGAACTGAGGCGCTTGT 1004
QY 243 CAAAGGTAATTAAGCATGCGCAAAACATTGAGGCTGAGGATTTGAACAAATGAACATCA 302
Db 1003 CGATGCTAGGCTGACGACACTGTTGATGACATTCAGAAAGTGATGCTGCTGAGGC 944
QY 303 AGTGATCTTAATGTCAAAAGTGAAGAGGTAATTGTAAAGCTCAATTTGTGATCGGTGT 362
Db 943 TGACCTACTGCTGTGATGCTGACTCAGACGCACTGCTGAGCGTGAAGCCGAACTTGT 884
QY 363 TCAGATGATATCTGCTGATGAGATATGATTTACATACAAATTTGGGATCTTCATCC 422
Db 883 TGACCTATGCTGAGGCGCTAGTACTGCTGAAAGCGGACGACCTGTTGATGCCGATTTC 824
QY 423 AACCACTCATGCTATTTGCGATATTCAGAAATTTTGTGCTTGTCCCTTGAATTTTC 482
Db 823 AGAGCCTGAGATCTTGTGAGGCTGAAGCACTTGTGCTGTAAGCCGATGCACTTGT 764
QY 483 TGATGAAGT 492
Db 763 TGATGCTGAT 754

RESULT 7

US-08-961-527-71/c
; Sequence 71, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 71:

SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-08-961-527-71

Query Match 5.2%; Score 39.6; DB 4; Length 32768;
Best Local Similarity 43.3%; Pred. No. 0.55;
Matches 186; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 63 TGTCCCGTGGCATTTTGTCCGCTTTCGCTGATCAATTCATATGATTAATAATCACCGA 122
Db 541 TGAACGGACGACACTGTTGATGCTGAGGCTAGAGGCTAGTACTTCTGAGGCTGAGGC 482
QY 123 AGAAATCAACAAAGCTAATGATGATGATGATGCTATTTGAACATCCGAAACATAGA 182
Db 481 ACTGTTGATGCCGATTCAGATGACACTGTTGAGCCGACGCACTTGTGACGCTGAGGC 422
QY 183 TCCATGAAAGTACCTGATCATGCGGATAAATTGCAAGCTCATGTTGGATTGTGGATT 242
Db 421 TGAACCACTGATCTTGTGAGGCTGACGCACTGCTTGTGTAAGCTGAGGCGCTTGT 362
QY 243 CAAAGGTAATTAAGCATGCGCAAAACATTGAGGCTGAGGATTTGAACAAATGAACATCA 302
Db 361 CGATGCTAGGCTGACGACACTGTTGATGACATTCAGAAAGTGATGCTTGTGAGGC 302
QY 303 AGTGATCTTAATGTCAAAAGTGAAGAGGTAATTGTAAAGCTCAATTTGTGATCGGTGT 362
Db 301 TGACCTACTGCTGTGATGCTGACTCAGACGCACTGCTGAGCGTGAAGCCGAACTTGT 242
QY 363 TCAGATGATATCTGCTGATGAGATATGATTTACATACAAATTTGGGATCTTCATCC 422
Db 241 TGACCTATGCTGAGGCGCTAGTACTGCTGAAAGCGGACGACCTGTTGATGCCGATTTC 182
QY 423 AACCACTCATGCTATTTGCGATATTCAGAAATTTTGTGCTTGTCCCTTGAATTTTC 482
Db 181 AGAGCCTGAGATCTTGTGAGGCTGAAGCACTTGTGCTGTAAGCCGATGCACTTGT 122
QY 483 TGATGAAGT 492
Db 121 TGATGCTGAT 112

RESULT 8

US-09-134-001C-527/c
; Sequence 527, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 527

LENGTH: 276
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-527

Query Match 5.2%; Score 39.4; DB 4; Length 276;
Best Local Similarity 50.3%; Pred. No. 0.11;
Matches 97; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 566 TGGATCAATTTTGGCGTTTATCTGATGATGACCGCTATTTTCCAGACACCGTAC 625
Db 240 TTGAATCAATGATGATGATCTTTCTCAAGTATGCTCCTTTTGAAAGACGCTAC 181
QY 626 GTAAGGAATGACCAAGTATTTGGCAGCAGCATTTTAAACGTGAATTTGAAAAAATTAAC 685
Db 180 ATACCTTTAAATCCATGATATTTAGCAGCATTTTATTTATTTACGATTAACAGGATG 121
QY 686 CAATGACATCATTTTTCACACTGTACATCTATTTACATGACGACAAATAAATAAATTT 745
Db 120 CATCATTTATATTTTCTTCAAGTTCACTTCAATTAATAATTTTCACAAAATTTT 61
QY 746 TTATTTTATTTTC 758


```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,017
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/137,175
; FILING DATE: 26-OCT-1993
; APPLICATION NUMBER: PCT/US92/08972
; FILING DATE: 22-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: BARBOUR-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: Ip90
; INDIVIDUAL ISOLATE: Isolate from I. persulcatus from
; INDIVIDUAL ISOLATE: Soviet Union
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 125..949
; OTHER INFORMATION: /product= "Ospa"
; NAME/KEY: CDS
; LOCATION: 959..1843
; OTHER INFORMATION: /product= "Ospb"
; US-08-479-017-7

Query Match
Best Local Similarity 5.1%; Score 38.8; DB 3; Length 1958;
Matches 100; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 54 TTGTGATGTCGCGTCGATGTTGTCGCGTTCGCGTGCATGATCAATGATGATA 113
DB 985 TTTAGATTTAGCTTTAATAGCATGTGACAAAAAGTGCTGACCCAAACACATGATCA 1044
QY 114 AATCACGAGAAGATCAACAAAGCTATTGATGATGCCATTGCTGCTATTGAACATCCGA 173
DB 1045 AGACGTTGAAGACTTAAAAAAGATCAAAAAGACGACTTAAAAAGATCTTCTTGGT 1104
QY 174 AACATATAGTCCAAATCAAGTACATGATGCGGCTAAATTCGAACGTCATGTTGTA 233
DB 1105 AACAGAGACAGCGTGAAGCTTTAATTAATCAATGAATTTTCATCGACGAAAAA 1164
QY 234 TGTGATTTCAAGGTGATTA 255
DB 1165 TGAAGACGATTAATATGATTA 1186

RESULT 12
; US-09-134-001C-1021
; Sequence 1021, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
```

```
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1021
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-1021

Query Match
Best Local Similarity 5.0%; Score 38.4; DB 4; Length 1170;
Matches 93; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 13 ATTAACAATCAAAAAAATCTTACAAATGATGAATTTTGTGATGTCGCGTG 72
DB 493 ATCGCATCAAAATCTTATTATCAATATGATTAAATGTTGATGAGCATGTGCTGT 552
QY 73 GCATTTGCGCGTTCGCGTGCATGATCAATTCATGATGATTAATCAACGAGAAATCAAC 132
DB 553 GACGTGTGATGATGTTATTCTGATGCGTCAATATGATGTTCTATCTCACACCAAGGAG 612
QY 133 AAGCTATTGATGATGCCATTCCTGCTATTGAACAATCCGAACAATAGATCAATGAAA 192
DB 613 AAGAAATTAGCGCCTGTGTTATTGATATCGTGAAGACTTAACGCAAGTTGCTTCTAT 672
QY 193 GTAC 196
DB 673 GAAC 676

RESULT 13
; US-09-342-459-1
; Sequence 1, Application US/09342459
; Patent No. 6251633
; GENERAL INFORMATION:
; APPLICANT: Yan, Kang
; APPLICANT: Pearce, Kenneth
; TITLE OF INVENTION: Flsa
; FILE REFERENCE: GM10229
; CURRENT APPLICATION NUMBER: US/09/342,459
; CURRENT FILING DATE: 1999-06-29
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-342-459-1

Query Match
Best Local Similarity 5.0%; Score 38.4; DB 4; Length 1413;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 12 AATAACAATCAAAAAAATCTTACAAATGATGAATTTTGTGATGTCGCGTG 71
DB 471 AATTCGATTTCAAAATTCATTTTATTAATGATTAATGCGTAGAAGCATGTGCTGT 530
QY 72 GCGATTTGCGCGTTCGCGTGCATGATCAATTCATGATGAATTAATACCGAAGAAATCA 131
DB 531 TGATGATTTAGATGTTACTGATGATCAATGATGATGATGATGATGATGATGATGAT 590
QY 132 CAAAGCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 191
DB 591 AAAAGCTTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 650
QY 192 AGTACCTGATCA 203
DB 651 TGAACCGGTGA 662

RESULT 14
; US-08-961-527-365/C
; Sequence 365, Application US/08961527
```

```
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 365:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1436 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-961-527-365
;
Query Match
Best Local Similarity 43.0%; Score 38; DB 4; Length 1436;
Matches 185; Conservative 0; Mismatches 245; Indels 0; Gaps 0;
;
QY 63 TGTGCGGTGATTTGTCGCGTGGTGGCGATTCATTCATGATAAATACCGA 122
   |||||
Db 765 TGAAGCGAGCAGCTGCTGATGCTGAGCGGAGCGCTGACTGCTGAGCGC 706
;
QY 123 AGAAATCAACAAAGCTTTGATGATGCCATTGCTATTGAACAATCCGAAACATAGA 182
   |||||
Db 705 ACTGTTGATGCCGATTCAGATGCACTGTTGAGGCGCGACGCTTGTGCGAGCGC 646
;
QY 183 TCCAAATGAAGTACGTATGATGCGGATTAATTCGAAGCTGATGTTGGTATTGTGATTT 242
   |||||
Db 645 TGAAGCAGCTGATCTGCTGAGCGCTGACGCGCTGACTTCTGAAGCTGAGCGCTTGT 586
;
QY 243 CAAAGTGAAATTAACATGCGAAACATTGAGGCTCGAGGATTTGAACAATGAACGCTCA 302
   |||||
Db 585 CGATGCTGAGGCTGACGACATGCTGTTGATGAGATTCAAGACTGATGCTTGAGGCG 526
;
QY 303 AGGTGATGCTATGTCGAAGGTGAAGGGTATTGTTAAACCTCATTTGTTGATCGGTGT 362
   |||||
Db 525 TGACGTACTGTTATGCTGACTGACGACGCTGTTGACGCTGAAGCGGAGGACCTTGT 466
;
QY 363 TCACGATGATTCGTCCTGATGATGATGATGATTTAGCATACAAATTTGGGATCTTCATCC 422
   |||||
Db 465 TGACGCTGATCTGAGCGCTGAGTACTGCTGAAGCGGACGACCTGTTGATGCCGATTC 406
;
QY 423 AACCACTGATGCTATTCGATTAATCAAGATTTGTTGTTGCTTCCCTTGAATTTTC 482
   |||||
Db 405 AGAGCGTGAATGATCTGCTGCAAGCTGAAGGCACTGTTGATGCTGAAGCTGATGACCTGT 346
;
QY 483 TGATGAAGT 492
   |||||
```

```
Db 345 TGATGCTGAT 336
;
RESULT 15
US-09-134-001C-2105
; Sequence 2105, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2105
; LENGTH: 1722
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
;
US-09-134-001C-2105
;
Query Match
Best Local Similarity 49.5%; Score 38; DB 4; Length 1722;
Matches 98; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
;
QY 263 GAACATTGAGGCTCGAGATTTGAACAATGAACGTCAGTGCATATGTCANAAG 322
   |||||
Db 41 GTAACTTTTGTAGTTTATTAAGAAAAAATTAAGATTTGTAATTTCCACTCGGTG 100
;
QY 323 GTGAAGAGGATTTGTTAAGCTCATTTGTTGATGCGTGTTCACGATGATTCGTCGA 382
   |||||
Db 101 GAGTAGCGGAATTCGCTAATAATATGTACATCGTTGAAGTAGAGGATGAATGTTATGT 160
;
QY 383 TGAATATGATTTAGCATACAAATTTGGGTGATCTTCATTCACACGATCATTTGCG 442
   |||||
Db 161 TGATGAGGATTTATGTTCCAGAGATGAATGCTGGGTGATGATGCTGTTCTCGT 220
;
QY 443 ATATTCAAGATTTGTTG 460
   |||||
Db 221 ACATTCAATATGTCATTG 238
;
RESULT 16
US-09-134-001C-438
; Sequence 438, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 438
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
;
US-09-134-001C-438
;
Query Match
Best Local Similarity 4.9%; Score 37.2; DB 4; Length 1776;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
;
QY 492 TAACATACATGACATCTTTTGAAGTAGCAGCAATTCGCTAATGTTGCAACCATATTTG 551
```

Db 510 TACCATTAATGCTTAAGAGATGACCAACAATATATCTCAAGATTTTAAGATATTTT 569
Oy 552 TGGCTTTCAATCTGATCCATTTTGGCTTTATCTGATGATGACCGCTATTTT 611
Db 570 CCTTTATCAGCCGTTAATTTTATTTGATGCTTTAGACTTACTTTAGAGCAATAGT 629
Oy 612 CCAAGACACCGTACGTAGGAAGATGACCAAGATTTGGCACCAGCATTTA 661
Db 630 CCACACGATGTAGCAATGACGATTTAAAGTATTTGAGCCACTATTTA 679

RESULT 17
US-09-004-838-94/c
Sequence 94, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Confering Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-0788100S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 7832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..7832
OTHER INFORMATION: /note= "RG2D continuation"
US-09-004-838-94

Query Match 4.9%; Score 37; DB 4; Length 7832;
Best Local Similarity 53.9%; Pred. No. 1.6;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Oy 107 ATGATTAATATCCGACCAATCAACAAGCTATGATGATGCGCATTCGCTATTGAC 166
Db 5503 ATTGAAAAAATAAAAAGAAATGTTATAGTACTTGATGATAAATGATTAATTAACAAT 5444
Oy 167 AATCGAACAATAGATCAATGAAGAAGTACTGATGATCGGATTAATTCGAACGTCATG 226
Db 5443 AAAAGTAACTGTTGTCAAATGAATGATTAATGATGCTCCCAATCCCAATCTCTGAG 5384

Oy 227 TTGCTATTGTGGATTTCAAAG 247
Db 5363 TTCGCATAGAAATTTGCAAG 5363

RESULT 18
US-08-242-932-7
Sequence 7, Application US/08242932
Patent No. 5595740
GENERAL INFORMATION:
APPLICANT: Brady, L. Jeannine
TITLE OF INVENTION: Cloning of No. 5595740-19A Fc Binding Forms of
TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,932
FILING DATE: 16-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: 0F142
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 932 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-242-932-7

Query Match 4.8%; Score 36.6; DB 1; Length 932;
Best Local Similarity 52.3%; Pred. No. 0.96;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Oy 85 GTTTCGGCTGATCCAAATTCATGATTAATAATCCAGAAATCAACAAAGCTATTGAT 144
Db 557 GATGTGATTAACACAAATTCATCTCATGACAGAAATAAGATTAAACAAATAATCAT 616
Oy 145 GATGCCATTGCTGCTATTGACAAATCCGAACAATAGATCCAAATGAAGTACCTATCAT 204
Db 617 GAACCAATGATTCGATGATGATTAAGAAATTAAGAAATTAACGAACCTAATAGA 676
Oy 205 GCCGATAATTCGACGTCATGTTGTTGGA 239
Db 677 CTGTTACACATCAACACATGAGAAGTTGAGAA 711

RESULT 19
US-08-714-481-7
Sequence 7, Application US/08714481
Patent No. 5766506
GENERAL INFORMATION:
APPLICANT: Brady, L. Jeannine
TITLE OF INVENTION: Cloning of No. 5766506-19A Fc Binding Forms of
TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 8

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Saliwanchik & Saliwanchik
;; STREET: 2421 N.W. 41st Street, Suite A-1
;; CITY: Gainesville
;; STATE: FL
;; COUNTRY: USA
;; ZIP: 32606
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/714,481
;; FILING DATE: 16-SEP-1996
;; CLASSIFICATION: 424
;;
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/242,932
;; FILING DATE: 16-MAY-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Saliwanchik, David R.
;; REGISTRATION NUMBER: 31,794
;; REFERENCE/DOCKET NUMBER: UF142
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 904-375-8100
;; TELEFAX: 904-372-5800
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 932 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-714-481-7

Query Match 4.8%; Score 36.6; DB 1; Length 932;
Best Local Similarity 52.3%; Pred. No. 0.96;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 85 GTTTCGGCTGATCCATTCCTATGATTAATCCGAGAAATCAACAAGCTATTGAT 144
DB 557 GATGTGGATTAACACAAATTCATCTCATGAACAGAAAATGAGTTTAAACAAAATTGAT 616
QY 145 GATGCCATTGCTGCTATTGAAACATCCGAAACATAGATCCAAATGAAAGTACCTGATCAT 204
DB 617 GAACAATAATGATTCGTGATGATATTAGAAATTAATTAATTAACGAACATAATAGA 676
QY 205 GCCGATTAATTCGAACGCTCATGTTGTTGTGGA 239
DB 677 CTGTTACACATCAACAAACATGAAGAGTTGAGAA 711

RESULT 20
PCT-US95-06111-7
;; Sequence 7, Application PC/US9506111
;; GENERAL INFORMATION:
;; APPLICANT: Faulmann, Ervin L.
;; APPLICANT: Street address: 186 Grinter Hall
;; APPLICANT: City: Gainesville Florida
;; APPLICANT: State/Province: US
;; APPLICANT: Country: Florida
;; APPLICANT: Postal code/zip: 32611
;; APPLICANT: Phone number: 904-392-8929
;; APPLICANT: Fax number: 904-392-6600
;; APPLICANT: Telex number:
;; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of
;; NUMBER OF INVENTION: the Group B Streptococcal Beta Antigens
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Saliwanchik & Saliwanchik
;; STREET: 2421 N.W. 41st Street, Suite A-1
;; CITY: Gainesville
;; STATE: FL

;; COUNTRY: USA
;; ZIP: 32606
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/06111
;; FILING DATE:
;; CLASSIFICATION:
;;
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/242,932
;; FILING DATE: 16-MAY-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Saliwanchik, David R.
;; REGISTRATION NUMBER: 31,794
;; REFERENCE/DOCKET NUMBER: UF142
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 904-375-8100
;; TELEFAX: 904-372-5800
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 932 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
PCT-US95-06111-7

Query Match 4.8%; Score 36.6; DB 5; Length 932;
Best Local Similarity 52.3%; Pred. No. 0.96;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 85 GTTTCGGCTGATCCATTCCTATGATTAATCCGAGAAATCAACAAGCTATTGAT 144
DB 557 GATGTGGATTAACACAAATTCATCTCATGAACAGAAAATGAGTTTAAACAAAATTGAT 616
QY 145 GATGCCATTGCTGCTATTGAAACATCCGAAACATAGATCCAAATGAAAGTACCTGATCAT 204
DB 617 GAACAATAATGATTCGTGATGATATTAGAAATTAATTAATTAACGAACATAATAGA 676
QY 205 GCCGATTAATTCGAACGCTCATGTTGTTGTGGA 239
DB 677 CTGTTACACATCAACAAACATGAAGAGTTGAGAA 711

RESULT 21
US-08-430-024-1
;; Sequence 1, Application US/08430024
;; Patent No. 5644030
;; GENERAL INFORMATION:
;; APPLICANT: Faulmann, Ervin L.
;; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
;; NUMBER OF INVENTION: AN IGA BINDING PROTEIN
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Saliwanchik & Saliwanchik
;; STREET: 2421 N.W. 41st Street
;; CITY: Gainesville
;; STATE: FL
;; COUNTRY: USA
;; ZIP: 32606
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/430,024
;; FILING DATE: 27-APR-1995
;; CLASSIFICATION: 530
;; PRIORITY APPLICATION DATA:

```

: APPLICATION NUMBER: US 07/677,209
: FILING DATE: 29-MAR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: BL-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2617 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus agalactiae
: STRAIN: DL471
: IMMEDIATE SOURCE:
: CLONE: pELF26
:
: FEATURE:
: NAME/KEY: sig-peptide
: LOCATION: 320..430
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 320..1510
: OTHER INFORMATION: /codon_start= 320
: OTHER INFORMATION: /function= "binds to Fc region of human IgA"
: OTHER INFORMATION: /product= "IgA binding protein"
: OTHER INFORMATION: /number= 1
:
: FEATURE:
: NAME/KEY: RBS
: LOCATION: 307..311
: FEATURE:
: NAME/KEY: protein_bind
: LOCATION: 887..1507
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /bound_molecule= "IgA Fc"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
: OTHER INFORMATION: /standard_name= "Human IgA-Fc binding"
: OTHER INFORMATION: /label= IgA-binding
:
: US-08-430-024-1
:
: Query Match 4.8%; Score 36.6; DB 1; Length 2617;
: Best Local Similarity 52.3%; Pred. No. 1.4;
: Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
:
: Oy 85 GTTGGCGTGAATCCATTCATGATATAAATCACCAGAAATCAACAACCTATTGAT 144
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 680 GATGTGATTAACACATCTATCTCATGAAACAGAAAATGAGTTTAAACAAAATGTGAT 739
:
: Oy 145 GATGCCATTGCTGCTATTGAAACAATCCGAAACAATAGATCCAATGAAGTACCTGATCAT 204
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 740 GAAACAATGATTCGATGATCATTTAAGATTAAGAAATCAATTTAACAACCAATTAAG 799
:
: Oy 205 GCCGATAAATTCGAAGTCATGTTGGTATTTGGGA 239
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 800 CTGTACACATCAACAACATGAAAGATTGAGAA 834
:
: RESULT 22
: US-08-782-009-1
: Sequence 1, Application US/08782009
: Patent No. 5714334
: GENERAL INFORMATION:
: APPLICANT: Faulmann, Ervin L
: TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/782,009
: FILING DATE: 07-JAN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/430,024
: FILING DATE: 27-APR-1995
: APPLICATION NUMBER: US 07/677,209
: FILING DATE: 29-MAR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: BL-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2617 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus agalactiae
: STRAIN: DL471
: IMMEDIATE SOURCE:
: CLONE: pELF26
: FEATURE:
: NAME/KEY: sig-peptide
: LOCATION: 320..430
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 320..1510
: OTHER INFORMATION: /codon_start= 320
: OTHER INFORMATION: /function= "binds to Fc region of human IgA"
: OTHER INFORMATION: /product= "IgA binding protein"
: OTHER INFORMATION: /number= 1
:
: FEATURE:
: NAME/KEY: RBS
: LOCATION: 307..311
: FEATURE:
: NAME/KEY: protein_bind
: LOCATION: 887..1507
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /bound_molecule= "IgA Fc"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
: OTHER INFORMATION: /standard_name= "Human IgA-Fc binding"
: OTHER INFORMATION: /label= IgA-binding
:
: US-08-782-009-1
:
: Query Match 4.8%; Score 36.6; DB 1; Length 2617;
: Best Local Similarity 52.3%; Pred. No. 1.4;
: Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
:
: Oy 85 GTTGGCGTGAATCCATTCATGATATAAATCACCAGAAATCAACAACCTATTGAT 144
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 680 GATGTGATTAACACATCTATCTCATGAAACAGAAAATGAGTTTAAACAAAATGTGAT 739
:
: Oy 145 GATGCCATTGCTGCTATTGAAACAATCCGAAACAATAGATCCAATGAAGTACCTGATCAT 204

```

Db 740 GAAACAAATGATTCGATGCATATTAGATAATGAAATCAATTAACGAACATAATAGA 799
QY 205 GCCGATAATTCGACGTCATGTGTATTTGGA 239
Db 800 CTGTTACACATCAAAACACATGACGAGATTGAGAA 834

RESULT 23

US-09-017-302-1
Sequence 1, Application US/09017302
Patent No. 6075128

GENERAL INFORMATION:
APPLICANT: Faulmann, Ervin L
TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/017,302
FILING DATE:

CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/782,009
FILING DATE: 07-JAN-1997
APPLICATION NUMBER: US 08/430,024
FILING DATE: 27-APR-1995
APPLICATION NUMBER: US 07/677,209
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: BL-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2617 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus agalactiae
STRAIN: DL471
IMMEDIATE SOURCE:
CLONE: pELF26

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 320..430

FEATURE:
NAME/KEY: CDS
LOCATION: 320..1510
OTHER INFORMATION: /codon_start= 320
OTHER INFORMATION: /function= "binds to Fc region of human IgA"
OTHER INFORMATION: /product= "IgA binding protein"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: RBS

LOCATION: 307..311
FEATURE:
NAME/KEY: protein_bind
LOCATION: 887..1507
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /bound_moiety= "IgA Fc"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "Human IgA-Fc binding"
OTHER INFORMATION: /label= IgA-binding
US-09-017-302-1

Query Match 4.8%; Score 36.6; DB 3; Length 2617;
Best Local Similarity 52.3%; Pred. NO. 1.4;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 85 GTTGGCGTCGATCCATTCATCTATGATAATCACCAGAAATCAACAAGCTATTGAT 144
Db 680 GATGTGGATACACAAATTTCTATCTCATGACAGAAAAAGAGTTTAAACAAATTTGAT 739
QY 145 GATGCCATTGCTGCTATTGACAAATCCGAAACAAATAGATCCAAAGTACCTGATCAT 204
Db 740 GAAACAAATGATTCGATGCATATTAGATAATGAAATCAATTAACGAACATAATAGA 799
QY 205 GCCGATAATTCGACGTCATGTGTATTTGGA 239
Db 800 CTGTTACACATCAAAACACATGACGAGATTGAGAA 834

RESULT 24

US-08-923-992A-7
Sequence 7, Application US/08923992A
Patent No. 6280738

GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438, 0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3294 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS

Db 373 CTGTTACACATCAACACATGAGAGTTGAGAA 407

RESULT 27

US-08-242-932-8
Sequence 8, Application US/08242932

Patent No. 5595740

GENERAL INFORMATION:

APPLICANT: Brady, L. Jeannine

TITLE OF INVENTION: Cloning of No. 5595740-1GA Fc Binding Forms of

TITLE OF INVENTION: the Group B Streptococcal Beta Antigens

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/242,932

FILING DATE: 16-MAY-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF142

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3730 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-242-932-8

Query Match 4.8%; Score 36.6; DB 1; Length 3730;
Best Local Similarity 52.3%; Pred. No. 1.6; Mismatches 74; Indels 0; Gaps 0;
Matches 81; Conservative 0;

QY 85 GTTTCGCTGATCAATTCATATGATTAATGACCGAGAAATCAACAAGCTATTGAT 144

Db 490 GATGTGATACACAAATTCATCTCATGACAGAAAATGAGTTAAACAAAATTGAT 549

QY 145 GATGCCATTGCTCTATTGACAAATCCGAAACATATGATCAATGAAAGTACTGATCAT 204

Db 550 GAAACAAATGATTCATGATGATTTAGAAATTTAGAAAATCAATTTAAGAACTAATAGA 609

QY 205 GCCGATAAATTCGACGTCATGTTGGTATTGTGGA 239

Db 610 CTGTTACACATCAACACATGAGAGTTGAGAA 644

RESULT 28

US-08-714-481-8
Sequence 8, Application US/08714481

Patent No. 576606

GENERAL INFORMATION:

APPLICANT: Brady, L. Jeannine

TITLE OF INVENTION: Cloning of No. 576606-1GA Fc Binding Forms of

TITLE OF INVENTION: the Group B Streptococcal Beta Antigens

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/714,481

FILING DATE: 16-SEP-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/242,932

FILING DATE: 16-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF142

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3730 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-714-481-8

Query Match 4.8%; Score 36.6; DB 1; Length 3730;
Best Local Similarity 52.3%; Pred. No. 1.6; Mismatches 74; Indels 0; Gaps 0;
Matches 81; Conservative 0;

QY 85 GTTTCGCTGATCAATTCATATGATTAATGACCGAGAAATCAACAAGCTATTGAT 144

Db 490 GATGTGATACACAAATTCATCTCATGACAGAAAATGAGTTAAACAAAATTGAT 549

QY 145 GATGCCATTGCTCTATTGACAAATCCGAAACATATGATCAATGAAAGTACTGATCAT 204

Db 550 GAAACAAATGATTCATGATGATTTAGAAATTTAGAAAATCAATTTAAGAACTAATAGA 609

QY 205 GCCGATAAATTCGACGTCATGTTGGTATTGTGGA 239

Db 610 CTGTTACACATCAACACATGAGAGTTGAGAA 644

RESULT 29

PCT-US95-06111-8
Sequence 8, Application PCT/US9506111

GENERAL INFORMATION:

APPLICANT:

APPLICANT: Street address: 186 Ginter Hall

APPLICANT: City: Gainesville

APPLICANT: State/Province: Florida

APPLICANT: Country: US

APPLICANT: Postal code/Zip: 32611

APPLICANT: Phone number: 904-392-8929

APPLICANT: Telex number: 904-392-6600

TITLE OF INVENTION: Cloning of Non-1GA Fc Binding Forms of

TITLE OF INVENTION: the Group B Streptococcal Beta Antigens

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

```

:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/06111
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/242,932
: FILING DATE: 16-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: UP142
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3730 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: PCT-US95-06111-8
:
: Query Match      4.8%; Score 36.6; DB 5; Length 3730;
: Best Local Similarity 52.3%; Pred. No. 1.6;
: Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```

```

Oy 85 GTTTCGGCTGATCCAAATTCACATGATGATAAATCACCAGAAATGCAACAACTATTGAT 144
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 490 GATGTGGATTAACACAAATTCATCTCATGACAGAAAATGAGCTTTAAACAAAATTGAT 549
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 145 GATGCCATTGCTGCTATTGACCAATCCGAACATAGATCCAAATGAACTACCTGATCAT 204
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 550 GAAACAAATGATCTGATGATGATTAAGAAATTCATTTAAGAAATCAATTAACGAAACTAATAGA 609
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 205 GCCGATAATTGCAACGTCATGTTGCTATTGTGGA 239
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 610 CTGTTACATCAACACATGAGAGAGTGTGAAA 644
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

: RESULT 30
: US-08-242-932-1
: Sequence 1, Application US/08242932
: Patent No. 5595740
: GENERAL INFORMATION:
: APPLICANT: Brady, L. Jeannine
: TITLE OF INVENTION: Cloning of No. 5595740-1ga Fc Binding Forms of
: TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/242,932
: FILING DATE: 16-MAY-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794

```

```

:
: REFERENCE/DOCKET NUMBER: UP142
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4200 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-242-932-1
:
: Query Match      4.8%; Score 36.6; DB 1; Length 4200;
: Best Local Similarity 52.3%; Pred. No. 1.7;
: Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```

```

Oy 85 GTTTCGGCTGATCCAAATTCACATGATGATAAATCACCAGAAATGCAACAACTATTGAT 144
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 680 GATGTGGATTAACACAAATTCATCTCATGACAGAAAATGAGCTTTAAACAAAATTGAT 739
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 145 GATGCCATTGCTGCTATTGACCAATCCGAACATAGATCCAAATGAACTACCTGATCAT 204
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 740 GAAACAAATGATCTGATGATGATTAAGAAATTCATTTAAGAAATCAATTAACGAAACTAATAGA 799
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 205 GCCGATAATTGCAACGTCATGTTGCTATTGTGGA 239
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 800 CTGTTACATCAACACATGAGAGAGTGTGAAA 834
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

: RESULT 31
: US-08-714-481-1
: Sequence 1, Application US/08714481
: Patent No. 5766606
: GENERAL INFORMATION:
: APPLICANT: Brady, L. Jeannine
: TITLE OF INVENTION: Cloning of No. 5766606-1ga Fc Binding Forms of
: TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/714,481
: FILING DATE: 16-SEP-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/242,932
: FILING DATE: 16-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: UP142
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4200 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-714-481-1

```

Query Match 4.8%; Score 36.6; DB 1; Length 4200;
Best Local Similarity 52.3%; Pred. No. 1.7;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 85 GTTTCGGCTGATTCATTCCTATGATTAATAATACCGAAGAAATCAACAAGCTATTGAT 144
DB 680 GATGTGGATTAACAAATTCATCTCATGTAACAGAAAAATGAGTTTAAACAAAAATTGAT 739

QY 145 GATGCCATTGCTGCTATTTGAACAATCGAACAATAGATCCATGGAAGTACCTGATCAT 204
DB 740 GAACAACAATGATTCGTGATTTATTAAGAAATTTGAATAATTTAAGCAACTAATAGA 799

QY 205 GCCGATAAATTCGACGTCATGTTGGTATTGTGA 239
DB 800 CTGTTACACATCAACAACATGAAGAAGTTGAGAA 834

RESULT 32
US-08-923-992A-1
; Sequence 1, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-19A Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 320..3811
; US-08-923-992A-1

Query Match 4.8%; Score 36.6; DB 4; Length 4200;
Best Local Similarity 52.3%; Pred. No. 1.7;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 85 GTTTCGGCTGATTCATTCCTATGATTAATAATACCGAAGAAATCAACAAGCTATTGAT 144
DB 680 GATGTGGATTAACAAATTCATCTCATGTAACAGAAAAATGAGTTTAAACAAAAATTGAT 739

QY 145 GATGCCATTGCTGCTATTTGAACAATCGAACAATAGATCCATGGAAGTACCTGATCAT 204
DB 740 GAACAACAATGATTCGTGATTTATTAAGAAATTTGAATAATTTAAGCAACTAATAGA 799

QY 205 GCCGATAAATTCGACGTCATGTTGGTATTGTGA 239
DB 800 CTGTTACACATCAACAACATGAAGAAGTTGAGAA 834

RESULT 33
PCT-US95-06111-1
; Sequence 1, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville
; APPLICANT: State/Province: Florida
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 32611
; APPLICANT: Phone number: 904-392-8929
; APPLICANT: Fax number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-19A Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US95-06111-1

Query Match 4.8%; Score 36.6; DB 5; Length 4200;
Best Local Similarity 52.3%; Pred. No. 1.7;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 85 GTTTCGGCTGATTCATTCCTATGATTAATAATACCGAAGAAATCAACAAGCTATTGAT 144
DB 680 GATGTGGATTAACAAATTCATCTCATGTAACAGAAAAATGAGTTTAAACAAAAATTGAT 739

QY 145 GATGCCATTGCTGCTATTTGAACAATCGAACAATAGATCCATGGAAGTACCTGATCAT 204
DB 740 GAACAACAATGATTCGTGATTTATTAAGAAATTTGAATAATTTAAGCAACTAATAGA 799

Qy 205 GCCGATTAATTCGACGTCATGTTGTAATGTCGA 239
| | | | | | | | | | | | | | | | | | | | | |
Db 800 CTGTACACATCAACACATGAAGAAGTTGACAA 834

RESULT 34

US-08-602-010A-1/C

; Sequence 1, Application US/08602010A

; Patent No. 5753235

; GENERAL INFORMATION:

; APPLICANT: Haanes, Elizabeth J.

; APPLICANT: Frank, Rexann S.

; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross & McIntosh

; STREET: 1700 Lincoln Street, Suite 3500

; CITY: Denver

; STATE: Colorado

; COUNTRY: U.S.A

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/602,010A

; FILING DATE: February 15, 1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Connell, Gary J.

; REGISTRATION NUMBER: 32,020

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303/863-9700

; TELEFAX: 303/863-0223

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5495 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-602-010A-1

Query Match 4.8%; Score 36.6; DB 1; Length 5495;

Best Local Similarity 54.0%; Pred. No. 1.8;

Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 614 AAGACACCGTACGTAAAGAAATGACCAAGTATGGCACCAGCATTTAAACGTGAATTGG 673
| | | | | | | | | | | | | | | | | | | | | |
Db 3716 AAACAGCGTATACCAACATGATATATTTAGCTAGATGTAATAAATAATTA 3657
| | | | | | | | | | | | | | | | | | | | | |
Qy 674 AAAAAATTACCAATGACATCATTTTCCAACTGACATCTCATTTCACTGACANT 733
| | | | | | | | | | | | | | | | | | | | | |
Db 3656 ATAAAAAGAGCTTCAACACCGCTTTATATATAAACATATAAATTCACGCCCAT 3597
| | | | | | | | | | | | | | | | | | | | | |
Qy 734 AAAATAAATTTTATTTT 752
| | | | | | | | | | | | | | | | | | | | | |
Db 3596 TTATGAATTTATTTGTTT 3578
| | | | | | | | | | | | | | | | | | | | | |

RESULT 35

US-08-602-010A-2

; Sequence 2, Application US/08602010A

; Patent No. 5753235

; GENERAL INFORMATION:

; APPLICANT: Haanes, Elizabeth J.

; APPLICANT: Frank, Rexann S.

; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A

ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/602,010A

; FILING DATE: February 15, 1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Connell, Gary J.

; REGISTRATION NUMBER: 32,020

; REFERENCE/DOCKET NUMBER: 2618-46

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303/863-9700

; TELEFAX: 303/863-0223

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5495 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-602-010A-2

Query Match 4.8%; Score 36.6; DB 1; Length 5495;

Best Local Similarity 54.0%; Pred. No. 1.8;

Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 614 AAGACACCGTACGTAAAGAAATGACCAAGTATGGCACCAGCATTTAAACGTGAATTGG 673
| | | | | | | | | | | | | | | | | | | | | |
Db 1780 AAACAGCGTATACCAACATGATATATTTAGCTAGATGTAATAAATAATTA 1839
| | | | | | | | | | | | | | | | | | | | | |
Qy 674 AAAAAATTACCAATGACATCATTTTCCAACTGACATCTCATTTCACTGACANT 733
| | | | | | | | | | | | | | | | | | | | | |
Db 1840 ATAAAAAGAGCTTCAACACCGCTTTATATATAAACATATAAATTCACGCCCAT 1899
| | | | | | | | | | | | | | | | | | | | | |
Qy 734 AAAATAAATTTTATTTT 752
| | | | | | | | | | | | | | | | | | | | | |
Db 1900 TTATGAATTTATTTGTTT 1918
| | | | | | | | | | | | | | | | | | | | | |

RESULT 36

US-08-680-726A-1/C

; Sequence 1, Application US/08680726A

; Patent No. 5804197

; GENERAL INFORMATION:

; APPLICANT: Haanes, Elizabeth J.

; APPLICANT: Frank, Rexann S.

; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES

; NUMBER OF SEQUENCES: 92

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross & McIntosh

; STREET: 1700 Lincoln Street, Suite 3500

; CITY: Denver

; STATE: Colorado

; COUNTRY: U.S.A.

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/680,726A

; FILING DATE: 12-JUL-1996

; CLASSIFICATION: 424

```
ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5495 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-680-726a-1

Query Match
Best Local Similarity 54.0%; Score 36.6; DB 1; Length 5495;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 614 AAGACACCGTACGTAGGAATGACCAAGATATTGGCAGCATTTAAAGCTGAATTGG 673
Db 3716 AAACACGCGTATACACATGATATATTTGTAGCTAGATGTAATAAATAATTA 3657
QY 674 AAAAAATTACCAATAGACATCATTTTCCAACTGTACCAATCTCTATTTCAGTGACAAT 733
Db 3656 ATAAAAAGAGCTTCAACACCGCTTTATATATAAACAAATAAAAAATTCCAGCCCATTT 3597
QY 734 AAATAAATTTTATTATT 752
Db 3596 TTTATGAATTTATTGTTT 3578

RESULT 37
US-08-680-726a-2
: Sequence 2, Application US/08680726a
: Patent No. 5804197
: GENERAL INFORMATION:
: APPLICANT: Haanes, Elizabeth J.
: APPLICANT: Frank, Rexann S.
: TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
: NUMBER OF SEQUENCES: 92
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/680,726a
: FILING DATE: 12-JUL-1996
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2618-46-C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5495 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-680-726a-2
```

```
Query Match
Best Local Similarity 54.0%; Score 36.6; DB 1; Length 5495;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 614 AAGACACCGTACGTAGGAATGACCAAGATATTGGCAGCATTTAAAGCTGAATTGG 673
Db 1780 AAACACGCGTATACACATGATATATTTGTAGCTAGATGTAATAAATAATTA 1839
QY 674 AAAAAATTACCAATAGACATCATTTTCCAACTGTACCAATCTCTATTTCAGTGACAAT 733
Db 1840 ATAAAAAGAGCTTCAACACCGCTTTATATATAAACAAATAAAAAATTCCAGCCCATTT 1899
QY 734 AAATAAATTTTATTATT 752
Db 1900 TTTATGAATTTATTGTTT 1918

RESULT 38
US-09-092-409-1/c
: Sequence 1, Application US/09092409
: Patent No. 6159478
: GENERAL INFORMATION:
: APPLICANT: Haanes, Elizabeth J.
: APPLICANT: Frank, Rexann S.
: TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
: NUMBER OF SEQUENCES: 92
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/092,409
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/680,726
: FILING DATE: 12-JUL-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2618-46-C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5495 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-09-092-409-1

Query Match
Best Local Similarity 54.0%; Score 36.6; DB 3; Length 5495;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 614 AAGACACCGTACGTAGGAATGACCAAGATATTGGCAGCATTTAAAGCTGAATTGG 673
Db 3716 AAACACGCGTATACACATGATATATTTGTAGCTAGATGTAATAAATAATTA 3657
QY 674 AAAAAATTACCAATAGACATCATTTTCCAACTGTACCAATCTCTATTTCAGTGACAAT 733
Db 3656 ATAAAAAGAGCTTCAACACCGCTTTATATATAAACAAATAAAAAATTCCAGCCCATTT 3597
```



```
Qy 614 AAGACACCGTAGTAAGAAATGACCAAGATATTGGCACCAGCATTTAAACGTGAATTGG 673
      || ||| ||| || || || ||| || || ||| || || ||| || || |||
Db 6886 AAAAAAGCGTATACAAACATGATATATATTTAGCTAGATCTATAAATAATTA 6945
Qy 674 AAAAAATTTAACCAATGACATCATTTTCCAACTGTACAACTCTATTCTACGTGCAAT 733
      | |||| | || || || || || || || || || || || || || || || ||
Db 6946 AATAAAAGAGCTTCCAAACCCGTTTATATATAAACAATAAATAATTCACGCCCAT 7005
Qy 734 AAAATATAATTTTATTTT 752
      || |||| || || || || || || || || || || || || || || || ||
Db 7006 TTTATGAATAATTATTTGTTT 7024

RESULT 44
US-09-134-001C-2243
; Sequence 2243, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2243
; LENGTH: 11091
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2243

Query Match 4.8%; Score 36.6; DB 4; Length 11091;
Best Local Similarity 56.1%; Pred. No. 2.3;
Matches 69; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 74 CATTTGTCGCCGTTTGGCTGATCCATTCATGATGAATAATCCAGCAAGATCAACA 133
      || ||| || || || || || || || || || || || || || || || || ||
Db 9636 CATTAAGTAATCATTCATTATTATTAACAATTAATTAATTAATCAAGAGTTTATG 9685
Qy 134 AAGCATATGATGATGATGATGCTGCTGCTATTGTAACATTCGAAACATAGATCAATGAAG 193
      | ||| || || || || || || || || || || || || || || || || || ||
Db 9686 ATACAAAGACAAACGATGAGCTATTCATTAATAATCAACCAATATCACTATCAAC 9745
Qy 194 TAC 196
      ||
Db 9746 CAC 9748

RESULT 45
US-08-961-527-373/C
; Sequence 373, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 373:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-373

Query Match 4.8%; Score 36.4; DB 4; Length 1068;
Best Local Similarity 51.2%; Pred. No. 1.1;
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 327 AGAGGCTATTGTTAAAGCTCATTTGTTGATCGGTTCACGATGATTCGATGGA 386
      || | || || || || || || || || || || || || || || || || || ||
Db 881 AGACGACCTCGTGGAGCGGAGCCGAAGCACTTGTGCTGATGCTGAGGCGCTGTGT 822
Qy 387 ATATGATTTAGCATCAATTAATGGTGATCTTCATCCACCACTCATGTCATTGCGATAT 446
      | || || || || || || || || || || || || || || || || || || ||
Db 821 ACTTGCTGAAGCGGACGACACTGCTTGATGCCATTCACAGCTGAGATCTTGAGAGC 762
Qy 447 TCAAGATTTTGTGTTGCCCTTGCCCTGGAATTTCTGATGAAGT 492
      || || || || || || || || || || || || || || || || || || || ||
Db 761 TGAACACTTGTGCTGCTGGAAGCCGATGACACTGTGTGATGCTGAT 716

RESULT 46
US-09-134-001C-737
; Sequence 737, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 737
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-737

Query Match 4.7%; Score 36; DB 4; Length 1596;
Best Local Similarity 44.4%; Pred. No. 1.7;
Matches 144; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

Qy 136 GCTATTGATGATGCCATTCGCTGCTATTGAAACAATCCGAACAATAGATCAATGAAGTA 195
      || ||| || || || || || || || || || || || || || || || || ||
Db 409 GCATGTGAAGGGAATCTTGTAGTATGCTGCCAAGATATGAAGCAACAACCTTA 468
Qy 196 CCGATCATGCCGGAATAATTTGAACGTCATGTTGGTATTGGAATTTAAAGGGAATTA 255
      || || || || || || || || || || || || || || || || || || || ||
Db 469 GCAAAAGTTTATTAGCAATTAACGATTTGGAATTTGGCACTTTGCTTAATTAATAA 528
Qy 256 GCCATGCAACATTTGAGGCTCGAGGATTTGAACAATAAGAAACGTCAGATGATCTAAT 315
```

Db 529 GACCTGCTGCACCTGACCGCCGATAGATTAAAGCAATTAAGATGTTATAGGATTA 588
Qy 316 GTCAAGGTGAAGAGGATATTGTTAAAGCTATTGTCGTCGTCAGATGATATC 375
Db 589 GATCAAGAGATATAGTACTGTCAGAGTCTAAGTCAATATAGTATGAGAAATTTTA 648
Qy 376 GTCTCGATGGAATATGATTAGCATCAAAATTTGGTGTGATCTTCATCCACCACTCATGTC 435
Db 649 GAGAAATATAGTATGATGTTGATCCAGCAGCGAGGTGATCCAGAGGCCCACTTAAGCA 708
Qy 436 ATTCGATATTCACATTTTGT 459
Db 709 CTATCTTGATTCAGAAATATGAT 732

RESULT 47
US-08-321-587-1
; Sequence 1, Application US/08321587
; Patent No. 5554526
; GENERAL INFORMATION:
; APPLICANT: Kunio KONDO et al.
; TITLE OF INVENTION: Human Parainfluenza Virus Type 4A Fusion
; TITLE OF INVENTION: Protein and Gene Coding for the Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/321,587
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/915,826
; FILING DATE: July 28, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2136 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:

LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-321-587-1

Query Match
Best Local Similarity 45.7%; Pred. No. 1.9;
Matches 126; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

Qy 414 TCTTCATCCACACCATGCTCATTTGCGATATTCAGATTTTGTGTCCTTGCCCT 473
Db 1556 TCTGCATCTTATTAATATGACAGTTACATTCGATTAATTTGGCTCAAGTATGACGA 1615
Qy 474 TGAATTTCTGATGAGGTACATACATGACATCTTTTGAAGTACGACATTCGCTGA 533
Db 1616 AGAAGTTAAGACGTGGCAGAAATCAAGACTTAATAGATGCTGATCTTTTAA 1675
Qy 534 TGTTCACACCATATGTTGCTCTTCATCTTGATCCATTTTGGCGTTTATCTGA 593
Db 1676 AGATCCCTAGTCAAAATACGCGTCTAGATTAATACGCCAAGATTCATATTTATAAC 1735
Qy 594 TGTATGACCGCTATTTTCCAAAGACCGTACGTAAAGAAATACCAAGTATTTGCACC 653
Db 1736 TTTATTTATCTAGATGCTCCAGAAAAATCATCCACAAAGATGTACCAATTCATCTCA 1795
Qy 654 AGCATTTAAAGTGAATTTGAAAAAATTAAACAT 689
Db 1796 TTGAAGAAACTCAAAATTCACAGATTCACACTAT 1831

RESULT 48
US-08-961-527-374/C
; Sequence 374, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373/7
REFERENCE/DOCKET NUMBER: PB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 374:
SEQUENCE CHARACTERISTICS:
LENGTH: 620 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

Query Match	4.78;	Score 35.8;	DB 4;	Length 620;
Best Local Similarity	45.48;	Pred. No. 1.4;		
Matches 124; Conservative	1;	Mismatches 148;	Indels 0;	Gaps 0;

OY	328	GAGGGATGTTGAAGCGCATATTGTGATGCGTGTACAGATGATTAATGCTGCATGGAA	387
Db	307	GACCACTCGTTGATGTGCTGAAGCCGGAAGCATTGTTCAGCTTGAAAGCGAAGCACTGGTT	248
OY	388	TATGATTTAGCATCAAAATTGGGTGATCTTCATGCCAACCACTCATGCAATTTGGATATT	447
Db	247	GACCTGTAARCGAAGCACACTGGTTGATGACAGATTACAGAAGCTGATGTGCTGCGAAGCT	188
OY	448	CMAATTTTGTGTGGCTTGTCCCTTGAAATTTCTGATGAAGTTAACTATPACATATACA	507
Db	187	GACCACCTGTTGATGCTGAAGACCGGAAGCACTCGTTATGCCGATTCAGAGCACTGTGC	128
OY	508	TCATTTGAAGTACACAAATTCGATAATGTGTCAACATATATGGTGTCTTCAATCTTG	567
Db	127	GATGCTGAAGCCGATGCACATGTACTGTGTAAACCGACGACACTGGGGCTTGCAGAGCC	68
OY	568	GATCCAATTTTGGCGTTTTATCTGATGATATG	600
		TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT	
		67 GAGGCACTGTTGAAGCTGATTCAGAGGCACTG	35

```

: RESULT 49
: US-09-485-077A-15
: Sequence 15, Application US/09485077A
: Patent No. 6458590
: GENERAL INFORMATION:
: APPLICANT: Mukherjee, Anil
: APPLICANT: Kundu, Gopal
: APPLICANT: Panda, Dibyendu
: TITLE OF INVENTION: Methods and Compositions for Treatment of Restenosis
: FILE REFERENCE: NIH-05047
: CURRENT APPLICATION NUMBER: US/09/485,077A
: CURRENT FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: PCT/US98/16569
: PRIOR FILING DATE: 1998-07-08
: PRIOR APPLICATION NUMBER: 60/054,967
: PRIOR FILING DATE: 1997-07-08
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patentin version 3.0
: SEQ ID NO 15
: LENGTH: 1422
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-485-077A-15

```

Query Match	4.78;	Score 35.4;	DB 4;	Length 1422;
Best Local Similarity	51.68;	Pred. No. 2.3;		
Matches 81; Conservative	0;	Mismatches 76;	Indels 0;	Gaps 0;
444	TATTCACAGATTTTGTGGCTTGCCCTGCTGCAATTTCTGATGAGGCTAACATACACT	503		

444 TATTCAAGATTTCGTTGCTTGCCCTTGCCCTTGAATTTCTGATGAAGGTACATAACAAT 503

[illegible]

```

1      RESULT 50
2      US-08-961-527-363/C
3      ; Sequence 363; Application US/08961527
4      ; Patent No. 6420135
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Charles Kunsch
7      ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
8      ; NUMBER OF SEQUENCES: 391
9      ; CORRESPONDENCE ADDRESS:
10     ; ADDRESSEE: Human Genome Sciences, Inc.
11     ; STREET: 9410 Key West Avenue
12     ; CITY: Rockville
13     ; STATE: Maryland
14     ; COUNTRY: USA
15     ;
16     ; ZIP: 20850
17     ;
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
20     ; COMPUTER: HP Vectra 486/33
21     ; OPERATING SYSTEM: MSDOS version 6.2
22     ; SOFTWARE: ASCII Text
23     ;
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/961,527
26     ;
27     ; FILING DATE:
28     ; CLASSIFICATION: 424
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER:
31     ;
32     ; FILING DATE:
33     ; ATTORNEY/AGENT INFORMATION:
34     ; NAME: Brookes, A. Anders
35     ; REGISTRATION NUMBER: 36,373
36     ; REFERENCE/DOCKET NUMBER: PB340P1
37     ;
38     ; TELECOMMUNICATION INFORMATION:
39     ; TELEPHONE: (301) 309-8504
40     ; TELEFAX: (301) 309-8512
41     ;
42     ; INFORMATION FOR SEQ ID NO: 363:
43     ; SEQUENCE CHARACTERISTICS:
44     ; LENGTH: 4483 base pairs
45     ; TYPE: nucleic acid
46     ; STRANDEDNESS: double
47     ;
48     ; TOPOLOGY: linear
49     ;
50     ; US-08-961-527-363

```

Query Match	4.68;	Score	35.2;	DB	4;	Length	4483;
Best Local Similarity	46.78;	Pred. No.	4;				
Matches 112; Conservative	0;	Mismatches	128;	IndeIs	0;	Gaps	0;

QY	360	TGTCACAGATGATATCGTCTCGANAGAAATATGATTTAGCATACAAATTGGGATCTTCA	413
Db	3687	TGTGCAGCGCTAGGCTACGCACTGGTACTCTGTGAAGCTGAGGCGCTGTTCGACGCTGA	3628
QY	420	TCCAACCACTCATCTCAATTTTCGGATATATCAAGATTTTGTGTGGCTTGTCCCTTGAAT	479
Db	3627	TTTCAGATGCTGATATGCTTGTCTGAGCCGAGGCACTTTTGCATGCTGAGGCTGAAAGCACT	3568
QY	480	TTTCTGATGAGAGTAAACATTAACAATTCATTTTGAAGTAGACAAATTCGCTAATGTTGT	539
Db	3567	TGTTTGATGCTGAGCACTGAACCACTGGTACTGTCTGAAGCTGACGCACTGGTTGACGCTGA	3508
QY	540	CAACCATATATGGTGGCTTTTCAATCTTGGATCCAAATTTTGGCGTTTATCTGATGATTT	599
Db	3507	AGCCGAGGCACTGGTTTATCTCTGAGCCGATGACATCGTTTATGCTGGAAGCTGAGTACACT	3448

Fri Feb 21 10:03:58 2003

us-10-024-955-6_1.rni

Page 24

Search completed: February 20, 2003, 21:22:42
Job time : 198 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 21:20:03 ; Search time 114 Seconds
(without alignments)
3399.959 Million cell updates/sec

Title: US-10-024-955-6

Perfect score: 761
Sequence: 1 GATCTATATCATATACAAAT.....ATTATTATTTTATTCGCC 761

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	761	100.0	761	9	US-10-024-955-6
2	518.6	68.1	812	9	US-10-024-955-1
3	41	5.4	484	9	US-09-876-348A-46
4	41	5.4	484	9	US-09-876-796A-46
5	40.6	5.3	11871	10	US-09-070-927A-171
6	38.4	5.0	343	10	US-09-815-242-3069
7	38.4	5.0	1194	10	US-09-815-242-4734
8	38.4	5.0	1413	10	US-09-855-399-1
9	36.8	4.8	484	9	US-09-876-348A-47
10	36.8	4.8	484	9	US-09-876-796A-47
11	36.8	4.8	640681	10	US-09-938-842A-3721
12	36.6	4.8	2000	9	US-09-815-242-4724
13	36.6	4.8	2523	10	US-09-815-242-8653
14	36.6	4.8	2526	10	US-09-815-242-8653
15	36	4.7	31517	10	US-09-070-927A-180
16	35.6	4.7	549	10	US-09-815-242-1682
17	35.6	4.7	1266	10	US-09-815-242-8722
18	35.4	4.7	506	10	US-09-895-828-54
19	35.4	4.7	557	9	US-09-510-662A-6

Result No.	Score	Query Match	Length	ID	Description
20	35.4	4.7	557	10	US-09-778-320-6
21	35.4	4.7	557	10	US-09-910-689-6
22	35.4	4.7	557	12	US-10-010-742-6
23	35.4	4.7	558	10	US-09-895-828-424
24	35.4	4.7	1424	10	US-09-206-576-1
25	35.4	4.7	1469	10	US-09-919-770-1
26	35.4	4.7	1959	10	US-09-864-761-4012
27	35.4	4.7	3741	10	US-09-803-286A-1
28	35.2	4.6	2000	9	US-09-938-842A-4181
29	35	4.6	366	9	US-09-854-133-233
30	35	4.6	366	10	US-09-738-973-233
31	35	4.6	830	10	US-09-939-980-236
32	35	4.6	2000	9	US-09-938-842A-5067
33	35	4.6	4178	10	US-09-070-927A-93
34	35	4.6	4641	10	US-09-925-300-686
35	34.8	4.6	1896	10	US-09-815-242-9170
36	34.8	4.6	2460	9	US-09-938-842A-2580
37	34.8	4.6	3143	9	US-09-954-531-156
38	34.8	4.6	3143	9	US-09-954-531-368
39	34.8	4.6	3143	10	US-09-964-824A-557
40	34.8	4.6	3143	10	US-09-964-824A-557
41	34.6	4.5	5361	9	US-09-742-096-2
42	34.6	4.5	6152	9	US-09-742-096-1
43	34.6	4.5	57130	10	US-09-835-081-3
44	34.2	4.5	624	10	US-09-974-300-6524
45	34.2	4.5	709	10	US-09-765-272-95

ALIGNMENTS

RESULT 1
US-10-024-955-6
Sequence 6, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 761 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 43..681
SEQUENCE DESCRIPTION: SEQ. ID NO: 6:
US-10-024-955-6

Query Match 100.0%; Score 761; DB 9; Length 761;
Best Local Similarity 100.0%; Pred. No. 2.3e-188;
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTTATATCAATTAACATCCAAAAACATATCTTACAAAATGATGAATTTTGTG 60
DB 1 GATCTTATATCAATTAACATCCAAAAACATATCTTACAAAATGATGAATTTTGTG 60
QY 61 ATTGCTGCGGTGCGATTTTGGCGGTTGGCGATGCCATTCATGATTAATAATCAC 120
DB 61 ATTGCTGCGGTGCGATTTTGGCGGTTGGCGATGCCATTCATGATTAATAATCAC 120
QY 121 GAAGAATCAACAAGCTATTGATGATGCCATTTGCTGCTATTGAACATCCGAACAATA 180
DB 121 GAAGAATCAACAAGCTATTGATGATGCCATTTGCTGCTATTGAACATCCGAACAATA 180
QY 181 GATCCAAATGAAGTACCTGATCCGATTAATTCGAACGTCATGTTGATTTGAT 240
DB 181 GATCCAAATGAAGTACCTGATCCGATTAATTCGAACGTCATGTTGATTTGAT 240
QY 241 TTCAAAAGTCAATTAAGCAGCAACATTTAGGCTCGAGATTTGAACAAATGAACGT 300
DB 241 TTCAAAAGTCAATTAAGCAGCAACATTTAGGCTCGAGATTTGAACAAATGAACGT 300
QY 301 CAAGTGATCTATGTCGAAGGTGAAGAGGATTTTAAAGCTCATTTGTTGATCGT 360
DB 301 CAAGTGATCTATGTCGAAGGTGAAGAGGATTTTAAAGCTCATTTGTTGATCGT 360
QY 361 GTTCACAGATATTCGTCGTGATGATTAATTTAGCATACAAATGGGTGATCTTCAT 420
DB 361 GTTCACAGATATTCGTCGTGATGATTAATTTAGCATACAAATGGGTGATCTTCAT 420
QY 421 CCACACCTCATGTCATTTGGGATTCAGATTTTGTGCTGCTGCTGCTGGAAT 480
DB 421 CCACACCTCATGTCATTTGGGATTCAGATTTTGTGCTGCTGCTGCTGGAAT 480
QY 481 TCTGATGAAGTAACATTAACATGACATCTTTGAATAGACAAATTCGTAATGTTGC 540
DB 481 TCTGATGAAGTAACATTAACATGACATCTTTGAATAGACAAATTCGTAATGTTGC 540
QY 541 AACCATATTGGGTCTTCAATCTTGATCAATTTTGGCGTTTATCTGATGAT 600
DB 541 AACCATATTGGGTCTTCAATCTTGATCAATTTTGGCGTTTATCTGATGAT 600
QY 601 ACCGCTATTTCCAGACGCGTACGTAAGAAATGACCAAGATTTGGACACGACT 660
DB 601 ACCGCTATTTCCAGACGCGTACGTAAGAAATGACCAAGATTTGGACACGACT 660
QY 661 AAACGTGAATTTGAAAAAATTAACCAATAGACATCTTTTCCAACTGTAACATCTCTA 720
DB 661 AAACGTGAATTTGAAAAAATTAACCAATAGACATCTTTTCCAACTGTAACATCTCTA 720
QY 721 TTTCACTGACATTAATAATTTTATTTTATTTCTCC 761
DB 721 TTTCACTGACATTAATAATTTTATTTTATTTCTCC 761

RESULT 2
US-10-024-955-1
; Sequence 1, Application US/10024955
; Patent No. US20020168373A1

GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kav-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
House Dust Mite and Uses Therefor

NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,36A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 812 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 68..712
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 119..712
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-024-955-1
Query Match 68.1%; Score 518.6; DB 9; Length 812;
Best Local Similarity 81.4%; Pred. No. 2.2e-125;
Matches 613; Conservative 0; Mismatches 139; Indels 1; Gaps 1;
QY 3 TCTTATATCAATTAACATCCAAAAACATATCTTACAAAATGATGAATTTTGTGAT 62
DB 28 TTTTTCATATCGTAATAAATCCAAATTCATCTTTTTCACAAATGATGAATTTATTTGAT 87
QY 63 TGTGCGCGTGCGATTTGTGCGGCTTGCGGTGATCCAAATTCATGATGAATTCACCGA 122
DB 88 TGTGCGCGTGCGATTTGTGCGGCTTGCGGTGATCCAAATTCATGATGAATTCACCGA 147
QY 123 AGAATCAACAAGCTATTGATGATGCCATTTGCTGATTTGAACAATCCGAACAATAGA 182
DB 148 AGAATTAACAAGCTGTTGATGAAGCGTGTGCAATTTGAAAAATCCGAACAATTCGA 207
QY 183 TCCAATGAAGAATCGATGATCCGATTAATTCGAACGTCATGTTGATTTGATTT 242
DB 208 TCCAATGAAGAATCCGATCATTTCTGATTAATTCGAACGATATTTGATTCATGATTT 267
QY 243 CAAGGTGAATTTAGCATTCGGAACATTTGAGGCTCGAGATTTGAACAATGAACGTCA 302
DB 268 AAAAGGTGAATTTAGCATTCGGAACATTTCAAGTTTCAGAGATTTAAACAATGAACGT 327
QY 303 AGGTGATGCTAATGCTAAAGGTGAAGAGGTATTTAAAGCTCATTTGTTGATTCGCTGT 362
DB 328 AGGTGATGCTAATGCTAAAGGTGAAGAGGTGTTTCAAGGCTCATTTGTTGCTGCTGT 387

```
Oy 363 TCACGATATATCGTCGATGGAATATGATTTAGCAACAATTTGGGATCTTCATCC 422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 TCATATGACGTTGTTTCATGATATGATTTAGCAACAATTTGGGATCTTCATCC 447
Oy 423 AACCACTATGTCATTTGGGATATTCAGATTTTGTGCTTGGCTTCCCTTGAATTC 482
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 448 AACCACTATGTCATTTGGGATATTCAGATTTTGTGCTTGGCTTCCCTTGAATTC 507
Oy 443 TGATGAGTAACTAACAAGACATCTTTGTAAGTACGACAATTCGTAATGTGTCAA 542
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 CGAAGAGTAAATATGATGATGACATCGTTCGAAAGTACGTAATTTGCCAAAGTGTCAA 567
Oy 543 CCATATTTGGTCTTTCATCTTGCATTCATTTTGGCGTTTATCTGATGATTTGAC 602
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 568 TCATATTTGGTCTTTCATCTTGCATTCATTTTGGCGTTTATCTGATGATTTGAC 627
Oy 603 CGCATTTTCCAAAGACCGTCAGTAAAGAAATGACCAAGATTTGGACACGACATTTAA 662
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 628 CGCATTTTCCAGATGACCGTCAGTAAAGAAATGACCAAGATTTGGACACGACATTTAA 687
Oy 663 ACGTAAATTTGCAAAAAATTAACAATAGACATCTTTTCCAACTGTACATCTCTATT 722
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 AAAAGAAATTTGCAAGAAA-CAACCAATGACTTACACACACACATTAACCTGTTATT 746
Oy 723 TCACGACATTAATAAATTTTATTTTAT 755
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 CACTGATTAATCAAAATTAATTTTAT 779
```

RESULT 3

```
US-09-876-348A-46
: Sequence 46, Application US/09876348A
: Patent No. US20020172951A1
: GENERAL INFORMATION:
: APPLICANT: Horwath, K. L. and Myers, K. L.
: TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
: TITLE OF INVENTION: Antifreeze Proteins and Method for Assaying Activity.
: FILE REFERENCE: RB-125-RI
: CURRENT APPLICATION NUMBER: US/09/876,348A
: CURRENT FILING DATE: 2001-06-09
: PRIOR APPLICATION NUMBER: 60/210,446
: PRIOR FILING DATE: 2000-06-08
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: Microsoft Word
: SEQ ID NO 46
: LENGTH: 484
: TYPE: DNA
: ORGANISM: Tenebrio molitor
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION:
: OTHER INFORMATION: Consensus of Seq ID #45 with B1/B2, 'n' defined as any nucleotide
US-09-876-348A-46
```

Query Match 5.4%; Score 41; DB 9; Length 484;

Best Local Similarity 16.9%; Pred. No. 0.19;

Matches 77; Conservative 69; Mismatches 310; Indels 0; Gaps 0;

```
Oy 52 TTTTGTGATTCGTCGCCGTCATTTGTCGCCGTCGATCAATTCATATGAT 111
| | | | | : : : : : | : | | | | |
Db 26 TTTTGTGATTCGTCGCCGTCATTTGTCGCCGTCGATCAATTCATATGAT 85
Oy 112 AAAATCACCAGAAATCAACAAGCTATGATGATGCTATGCTATTTGAACAATCC 171
| | | | | : : | | | : : | | | : : | | | : : | | |
Db 86 NAGNNANNNGNCCNARAYCAGCRNARARTGYNARNNNNGNNGAGGTGTCCNNAAGANRYN 145
Oy 172 GAAACAATAGATCAATGAAATGATCCGATTAATTTGAAGCTGATTTGGT 231
| | | | | : | | | | | : | | | | | : | | | | |
Db 146 ATNRNNNARAGYCCGANNNGTCNGNNGAYGAYCCYAAANTGAARNNNCANNTTYTY 205
Oy 232 ATTGTGATTTCAAGGTAATTAAGCATGCGAACAATTTAGCGCTGAGATTTGAACA 291
| | | | | : | | | | | : | | | | | : | | | | |
Db 206 TGCNTNNYNARGRNNNYGRNNTRGYNRCNGAANCNGAGANRYNRNNGYNANRYRYTN 265
```

```
Oy 292 ATGAACGTCAGGATGCTATATGTCAAAGTGAAGGGATGTTAAAGCTCATTTG 351
| | | | | : : : : : | : | | | | |
Db 266 ARGRNNNANAGTANGNRRNNNNNNNNNNRRNNRRRRRRRRRRRRRRNNRRNNRN 325
Oy 352 TTGATCGGTGTTCCAGCATGATATGTCGATGGAATGATTTAGCATACAAATTTGGT 411
| | | | | : | | | | | : | | | | | : | | | | |
Db 326 NNNNNNNNNNGARRRNNNNNNNNRRNNNNNNNNNNRRNNNNNNNNNNRRNNNNNN 385
Oy 412 GATCTTCATCCACCACTCATGTCATTTGCGATATTCAGATTTTGTGCTTGTCC 471
| | | | | : | | | | | : | | | | | : | | | | |
Db 386 RNNANNARCCNRNNTTYTYNCRNRYTRNTRNNNNNNNNNNNNNNNNNNNNRRNTTYRAN 445
Oy 472 CTGAAATTTGATGAGGATACATACATGACA 507
| | | | | : | | | | | : | | | | | : | | | | |
Db 446 AATTAAGNNNTYTRNTRNNRRNAAAAA 481
```

RESULT 4

```
US-09-876-796A-46
: Sequence 46, Application US/09876796A
: Patent No. US2002017302A1
: GENERAL INFORMATION:
: APPLICANT: Horwath, K. L. and Easton, C. M.
: TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
: TITLE OF INVENTION: Antifreeze Proteins and Method for Assaying Activity.
: FILE REFERENCE: RB-125-SEQ
: CURRENT APPLICATION NUMBER: US/09/876,796A
: CURRENT FILING DATE: 2001-06-07
: PRIOR APPLICATION NUMBER: 60/210,446
: PRIOR FILING DATE: 2000-06-08
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: Microsoft Word
: SEQ ID NO 46
: LENGTH: 484
: TYPE: DNA
: ORGANISM: Tenebrio molitor
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION:
: OTHER INFORMATION: Consensus of Seq ID #45 with B1/B2, 'n' defined as any nucleot
US-09-876-796A-46
```

Query Match 5.4%; Score 41; DB 9; Length 484;

Best Local Similarity 16.9%; Pred. No. 0.19;

Matches 77; Conservative 69; Mismatches 310; Indels 0; Gaps 0;

```
Oy 52 TTTTGTGATTCGTCGCCGTCATTTGTCGCCGTCGATCAATTCATATGAT 111
| | | | | : : : : : | : | | | | |
Db 26 TTTTGTGATTCGTCGCCGTCATTTGTCGCCGTCGATCAATTCATATGAT 85
Oy 112 AAAATCACCAGAAATCAACAAGCTATGATGATGCTATGCTATTTGAACAATCC 171
| | | | | : | | | | | : | | | | | : | | | | |
Db 86 NAGNNANNNGNCCNARAYCAGCRNARARTGYNARNNNNGNNGAGGTGTCCNNAAGANRYN 145
Oy 172 GAAACAATAGATCAATGAAATGATCCGATTAATTTGAAGCTGATTTGGT 231
| | | | | : | | | | | : | | | | | : | | | | |
Db 146 ATNRNNNARAGYCCGANNNGTCNGNNGAYGAYCCYAAANTGAARNNNCANNTTYTY 205
Oy 232 ATTGTGATTTCAAGGTAATTAAGCATGCGAACAATTTAGCGCTGAGATTTGAACA 291
| | | | | : | | | | | : | | | | | : | | | | |
Db 266 ARGRNNNANAGTANGNRRNNNNNNNNNNRRNNRRRRRRRRRRRRRRNNRRNNRN 325
Oy 352 TTGATCGGTGTTCCAGCATGATATGTCGATGGAATGATTTAGCATACAAATTTGGT 411
| | | | | : | | | | | : | | | | | : | | | | |
Db 326 NNNNNNNNNNGARRRNNNNNNNNRRNNNNNNNNNNRRNNNNNNNNNNRRNNNNNN 385
Oy 412 GATCTTCATCCACCACTCATGTCATTTGCGATATTCAGATTTTGTGCTTGTCC 471
| | | | | : | | | | | : | | | | | : | | | | |
```



```

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4734
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4734

Query Match
Best Local Similarity 50.0%; Score 38.4; DB 10; Length 1194;
Pred. No. 1.3;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
```

```

Qy 12 AATACAAATCCAAAAAACAATATCTTACAAATGAATTTTGTGATGCTGCCGT 71
Db 471 AATTCCTATTCAAAAATCGATTTTAATTAATGATTAATCGTAGAAGCATGTGGTGT 530
Qy 72 GGCATTTGTGCCGCTTTCGGCTGATCAATTCATGATGATAAATCAGCAAGAAATCA 131
Db 531 TGATGATTAAGATGTTTACTGATGCAATATGATGTTAATCCTAAGCTACTACTA 590
Qy 132 CAAGCTATGATGATGCCATGCTGCTATTGACAAATCCGAACAAATATCAATGAA 191
Db 591 AAAAGATTAGGTGATGCTGATGATGATGATGTAAGACGTTACGACAGTTGCTTTT 650
Qy 192 AGTACCTGATCA 203
Db 651 TGAACGCGGTGA 662

RESULT 8
US-09-855-399-1
; Sequence 1, Application US/09855399
; Patent No. US20020048788A1
; GENERAL INFORMATION:
; APPLICANT: Yan, Kang
; APPLICANT: Pearce, Kenneth
; TITLE OF INVENTION: ftsA
; FILE REFERENCE: GM10229
; CURRENT APPLICATION NUMBER: US/09/855,399
; CURRENT FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 09/342,459
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-855-399-1
```

```

Query Match
Best Local Similarity 50.0%; Score 38.4; DB 10; Length 1413;
Pred. No. 1.4;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
```

```

Qy 12 AATACAAATCCAAAAAACAATATCTTACAAATGAATTTTGTGATGCTGCCGT 71
Db 471 AATTCCTATTCAAAAATCGATTTTAATTAATGATTAATCGTAGAAGCATGTGGTGT 530
Qy 72 GGCATTTGTGCCGCTTTCGGCTGATCAATTCATGATGATAAATCAGCAAGAAATCA 131
Db 531 TGATGATTAAGATGTTTACTGATGCAATATGATGTTAATCCTAAGCTACTACTA 590
Qy 132 CAAGCTATGATGATGCCATGCTGCTATTGACAAATCCGAACAAATATCAATGAA 191
Db 591 AAAAGATTAGGTGATGCTGATGATGATGATGTAAGACGTTACGACAGTTGCTTTT 650
Qy 192 AGTACCTGATCA 203
Db 651 TGAACGCGGTGA 662
```

```

RESULT 9
US-09-876-348A-47
; Sequence 47, Application US/09876348A
; Patent No. US20020172951A1
; GENERAL INFORMATION:
; APPLICANT: Horwath, K. L. and Myers, K. L.
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; FILE REFERENCE: RB-125-RI
; CURRENT APPLICATION NUMBER: US/09/876,348A
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/210,446
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Word
; SEQ ID NO 47
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Tenebrio molitor
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION:
; OTHER INFORMATION: Consensus of SEQ. ID #46 with AFP-3, 'n' defined as any nucleo
US-09-876-348A-47
```

```

Query Match
Best Local Similarity 12.3%; Score 36.8; DB 9; Length 484;
Pred. No. 2.3;
Matches 56; Conservative 65; Mismatches 335; Indels 0; Gaps 0;
```

```

Qy 52 TTTTGTGATGATGCTGCCGCTGATGCTGCCGCTGATGCTGCCGCTGATGCTGCCGCT 111
Db 26 YNTGYTNRVYNYNYRYNYRYNTNTNRRCNNRRYVYANGCVNTNACYNANRANNNNNN 85
Qy 112 AAAATCACCAGAAATCAACAAGCTATTGATGATGCCATTGCTGCTATTGACAAATCC 171
Db 86 NAGNNRNNNNNNYNNRNNCAGCRNNRNNNTGYNRNNNNNNNGAGTPTCNAAAGANNYN 145
Qy 172 GAAACAATAGATCCAAATGAAGTACCTGATCATGCCGATTAATTTGCAAGCTCATTTGGT 231
Db 146 NTRNNNARRGYTCGCAANNNGTGNNNNGNGAYGATCCYAAANTGAARNNNCANNYYTY 205
Qy 232 ATTTGATTTCAAGGATGAATTAAGCAATGCGCAAAATTCGAGCTCGAGATTGAACAA 291
Db 206 TGCNTNNTNANRNNNNYGRNNTNRYNNNNNGNNNGGANNYNNNNNNYNNNNNTN 265
Qy 292 ATGAAAGCTCAAGGTGATGCTAATGTCAAAGGTGAAGGGTATTGTTAAAGCTCATTTG 351
Db 266 AARRNNNAARNTNANGNRRNNNNNNNNNNRNNRNNRNNRNNRNNRNNRNNRNNRNN 325
Qy 352 TTGATCGGTGTTCAAGGATGATATGCTGCGATGGAATGATTTGATTTGATTTGATTT 411
Db 326 NNNNNNNNNNGARRNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 385
Qy 412 GATCTTCATCCACCACTCATGCTCATTTGCGATATTCAAGATTTTGTGTTGCCCTTGTCC 471
Db 412 GATCTTCATCCACCACTCATGCTCATTTGCGATATTCAAGATTTTGTGTTGCCCTTGTCC 471
```

```
Db 386 RRANNAARNYCNNTNNNNNNNNNNYRNNTRRNNNNNNNNNNNNNNNNNNNNNNNNNNNN 445
QY 472 CTGGAATTTCGTGATGAAGGTAAACATACATGACA 507
Db 446 AATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 481

RESULT 10
US-09-876-796A-47
; Sequence 47, Application US/09876796A
; Patent No. US20020173024A1
; GENERAL INFORMATION:
; APPLICANT: Horwath, K. L. and Easton, C. M.
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding Type III Tenebrio
; FILE REFERENCE: RB-125-SEQ
; CURRENT APPLICATION NUMBER: US/09/876,796A
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/210,446
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Word
; SEQ ID NO 47
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Tenebrio molitor
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: Consensus of SEQ. ID #46 with AFP-3, 'n' defined as any nucleotide
US-09-876-796A-47 {

Query Match
Best Local Similarity 12.3%; Pred. No. 2.3; Length 484;
Matches 56; Conservative 65; Mismatches 335; Indels 0; Gaps 0;

QY 52 TTTTGTGATGTCGCGCGTCGTCATTTGTCGCCGTTGCGCTGATCCATTCATGAT 111
Db 26 YNNTGYTNNYNNYNNYRYNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 85
QY 112 AAATCCAGCAAGAAATCAACAAGCATATGATGATGCCATTTGCTGTATTTGAACATCC 171
Db 86 NAGNNRRNGNNNNYNNNNNCCACCRNNNNNTGTNARNNNNNNNNNNYGGAGRTCNNAAGANNYN 145
QY 172 GAAACATATAGATCCAAATGAAGTACCTGATCATGCCGATAAATTCGAACGTATGTTGT 231
Db 146 NTRNNRNNARBYCGCANNNCTGNNNNNGNGAYCGCYAANTGAARNNNCANNYYTY 205
QY 232 ATTGTGATTTCAAGGTGAATTAGCCATCGGAACATTGAGGCTCGAGGATTTGAACAA 291
Db 206 TGCNTNNYNAGRNNNNYGGHNTNRYNNRNNNNNNNNNGCANNYNNNNNNYNGANNNTN 265
QY 292 ATGAACGTCGAAGGTGATGCTAATGTCAAGGTGAAGAGGATATTGTTAAAGTCATTG 351
Db 266 ARRRNNNAARTNANGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 325
QY 352 TTGATCGGTTCGACGATGATATGCTCTCGATGGAATATGATTAGCATCAAAATGGT 411
Db 326 NNNNNNNNNNGARNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 385
QY 412 GATCTTCATCCACACACTCATGTCATTTGCGATATTTCAAGATTTTGTGTCCTTC 471
Db 386 RRANNAARNYCNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 445
QY 472 CTGGAATTTCGTGATGAAGGTAAACATACATGACA 507
Db 446 AATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 481

RESULT 11
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
```

```
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match
Best Local Similarity 4.8%; Score 36.8; DB 10; Length 640681;
Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 92 CTGATCCATTCACACTGATTAANAATCCCGAGAAATCAACAAGCTATTGATGATGCCA 151
Db 377627 CTGTACAAATGAAAATAAACAGTACGTAAAGTAAAGAAATATATGATTAATATA 377686
QY 152 TTGCTGCTATTGAACATCCGAAACATAGATCCAAATGGAAGTACCTGATCATGCCGATA 211
Db 377687 ATGATTTTATTTTGAAGAAAAGAAAAGAAAAGTAAACGAGAGTTCAGATGAACAAACTA 377746
QY 212 AATTGCAAGCTCATGTTGTATTGTGATTTCAAGGTGAATTAGCCA 259
Db 377747 AAATGCAACGCTATTAAACTGAAGATTTATTTATGTGAAAATGTA 377794

RESULT 12
US-09-938-842A-3721
; Sequence 3721, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepis, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCDIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3721
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3721

Query Match
Best Local Similarity 4.8%; Score 36.6; DB 9; Length 2000;
Matches 126; Conservative 0; Mismatches 129; Indels 2; Gaps 1;

QY 505 ACACTTTTGAAGTACACATTCGTTAATGTTGTCACACCATATTGTTGTCCTTCAATC 564
Db 165 ACTGATATCAAGATGATACATCAACATCAAGTAAACCTTTGGATGTCCTTTCTC 224
QY 565 TTGATTCATTTTGGCGCTTTATCTGATGTTGACCGCTATTTTCCAAAGACCCGTA 624
```

```
Db 225 TTGTCTAGTAATAATTAAATTATATACCTACCGAATTTGTAAACATATTAACATTTGACATG 284
QY 625 CGTAAGGAATATGACCAAGATTTGGACACGATTTAAACGTGAATTGGAA--AAAAATT 682
Db 285 TGCATTTTAATCCATTAATGACGTCACTCTATTAATCCGTAAGTAATAGTATATATA 344
QY 683 AACCAATAGACATCAATTTTCCAACTGTACATCTCTAATTTTCAGTACATAAATAA 742
Db 345 CATGAAATGATCCATTAATACCATCTGATAAATTTTGTGAAGTGGCTAAAAATATAAC 404
QY 743 TTTTATTTTATTTCT 759
Db 405 AGTTTTTTTTTTTGT 421

RESULT 13
US-09-815-242-4724/c
; Sequence 4724, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4724
; LENGTH: 2523
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4724

Query Match 4.8%; Score 36.6; DB 10; Length 2523;
Best Local Similarity 47.9%; Pred. No. 5.2; Indels 0; Gaps 0;
Matches 105; Conservative 0; Mismatches 114;

QY 386 AATATGATTTAGACATAATTTGGGTGATCTTCATCCAACTCATTCATTTGGGATA 445
Db 571 AATGATCTTATCATTAACATTTGTCCTGTTGTGAGCTGCAACAGTATTCACGGCA 512
QY 446 TTCACATTTTGTGTCCTGCTCCCTTGAATTTCTGATGAAGGTAACATTAACATGA 505
Db 511 TGGGATTTAAAGTCTTGGTTTAATAGTGTGTTTAAAGTGAATTAACATCTTTTG 452
QY 506 CATCTTTGAAGTACACATTCGTAATGTGTCAACCATATTTGGTCTTCAATCT 565
Db 451 CTGTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 392
QY 566 TGGATCCAAATTTTGGCGTTTATCTGATGATTTGACCG 604
```

```
Db 391 TTGTACATTCGTAAGTTTTTGTAGTAGAATTAATTAATTAATTAATTAATTAATTA 353

RESULT 14
US-09-815-242-8653/c
; Sequence 8653, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8653
; LENGTH: 2526
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2526)
US-09-815-242-8653

Query Match 4.8%; Score 36.6; DB 10; Length 2526;
Best Local Similarity 47.9%; Pred. No. 5.2;
Matches 105; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 386 AATATGATTTAGACATAATTTGGGTGATCTTCATCCAACTCATTCATTTGGGATA 445
Db 571 AATGATCTTATCATTAACATTTGTCCTGTTGTGAGCTGCAACAGTATTCACGGCA 512
QY 446 TTCACATTTTGTGTCCTGCTCCCTTGAATTTCTGATGAAGGTAACATTAACATGA 505
Db 511 TGGGATTTAAAGTCTTGGTTTAATAGTGTGTTTAAAGTGAATTAACATCTTTTG 452
QY 506 CATCTTTGAAGTACACATTCGTAATGTGTCAACCATATTTGGTCTTCAATCT 565
Db 451 CTGTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 392
QY 566 TGGATCCAAATTTTGGCGTTTATCTGATGATTTGACCG 604
Db 391 TTGTACATTCGTAAGTTTTTGTAGTAGAATTAATTAATTAATTAATTAATTAATTA 353

RESULT 15
US-09-070-927A-180
; Sequence 180, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
```



```
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 8722
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-815-242-8722
```

```
Query Match          4.7%; Score 35.6; DB 10; Length 1266;
Best Local Similarity 45.2%; Pred. No. 7.1;
Matches 131; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
```

```
OY 291 AATGAACCTCAAGTGATGCTAAGTCAAGGTAAGGCGTATTGTTAAAGCTCATTT 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 27 AATAAATTAACGGGTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 86
OY 351 GTTGATCGGTGTTACGATGATATCGTCTCGATGGAATGATTTGATTCATACAAATGGG 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 87 GACAGCATCTTATTAAGCTTGATTAACCGAGCAAAATTAGTTAATGTTCCAGCTTTAAG 146
OY 411 TGATCTTCATCCACCATCTCATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 147 TGATGTAGAAACATTAATATGATTAACAACTTTAAATGCTGACGTTACATACAAAA 206
OY 471 CCTGAATTTCTGATGAAGTAACATAACATGACATCTTTTGAAGTACGACAAATTCGC 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 GGACGAAATGCTGTGTGCTGATGCAACAAGACTTAATGAAGGACCATATGA 266
OY 531 TATGTGTGCAACCATATTGGTGCTTTTCATCTTGATTCCAATTTTGG 580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 AATGTAGTAAATGCTGCAAGTATTTAGTTAGGAGCGCTTTTAG 316
```

```
RESULT 18
; Sequence 54, Application US/09895828
; Patent No. US20020099012A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Watanabe, Yoshihito
; APPLICANT: Carter, Darriek
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.539
; CURRENT APPLICATION NUMBER: US/09/895.828
; NUMBER OF SEQ ID NOS: 473
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 54
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 487, 490
; OTHER INFORMATION: n = A,T,C or G
US-09-895-828-54
```

```
Query Match          4.7%; Score 35.4; DB 10; Length 506;
Best Local Similarity 51.6%; Pred. No. 5.7;
Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
```

```
OY 444 TATCAAGATTTTGTGTGCTTGCCTTGCCCTTGAATTTCTGATGAAGGTACATACAT 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 292 TATTTTAATATTTGTTATCTCATGAATAGAAATTAATGTAAGCAACAAATTAATCT 351
OY 504 GACATCTTTGAAGTACGACAAATTCGCTAATGTTGCAACCAATATGCTGCTTTCAT 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 TTATCCCACTTAAAGAGAAATATACATTTATGTCACATATATCTTTGTTTTTAAG 411
OY 564 CTGATCCCAATTTTGGCGTTTATCTGATGATATG 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 412 TTAGGTATATTTTGTGTGATTAATCTTTTGCGTG 448
```

```
RESULT 19
US-09-510-662A-6/C
; Sequence 6, Application US/09510662A
; Patent No. US20020155125A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Dillon, Davin C.
; APPLICANT: Wang, Aijun
; APPLICANT: Day, Craig H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.491C1
; CURRENT APPLICATION NUMBER: US/09/510,662A
; CURRENT FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 6
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-510-662A-6
```

```
Query Match          4.7%; Score 35.4; DB 9; Length 557;
Best Local Similarity 51.6%; Pred. No. 5.7;
Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
```

```
OY 444 TATCAAGATTTTGTGTGCTTGCCTTGCCCTTGAATTTCTGATGAAGGTACATACAT 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 TATTTTAATATTTGTTATCTCATGAATAGAAATTTTGTGACGAACAAATAATCT 126
OY 504 GACATCTTTGAAGTACGACAAATTCGCTAATGTTGCAACCAATATGCTGCTTTCAT 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 TTATCCCACTTAAAGAGAAATATACATTTATGTCACATATATCTTTGTTTTTAAG 66
OY 564 CTGATCCCAATTTTGGCGTTTATCTGATGATATG 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 TTAGGTATATTTTGTGTGATTAATCTTTTGCGTG 29
```

```
RESULT 20
US-09-778-320-6/C
; Sequence 6, Application US/09778320
; Patent No. US20010034052A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Day, Craig H.
; APPLICANT: Jiang, Yujin
```



```
RESULT 29
US-09-854-133-233/c
; Sequence 233, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 233
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(366)
; OTHER INFORMATION: n = A,T,C or G
US-09-854-133-233
```

```
Query Match          4.6%; Score 35; DB 9; Length 366;
Best Local Similarity 51.0%; Pred. No. 6;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

```
Oy 444 TATTCAAGATTGTTGTTGCTTGCCTTGACCTTGAATTTCTGATGAAGTAACATAACAT 503
    ||||| | ||||| || | ||||| || ||||| ||||| ||||| ||||| |||||
Db 263 TATTTAATATTGTTATTTATTTATTCATGATAGAAATTTATGTAGAACAAACAAATACT 204
Oy 504 GACATCTTTTGAAGTACGACGATTCGCTAATGTTGTCAACCATATTGTCCTTTCAAT 563
    ||||| | ||||| || | ||||| || ||||| ||||| ||||| ||||| |||||
Db 203 TTATCCCACTTAAAGAAATATTAACATTTATGTACATATAATCTTTGTTTTTAAG 144
Oy 564 CTTCGATCCCAATTTTGGCGTTTATCTGATGATGATG 600
    ||||| | ||||| || | ||||| || ||||| ||||| ||||| ||||| |||||
Db 143 TTAGTGTAATTTTGTGTGATTAATCTTTTGTGTG 107
```

```
RESULT 30
US-09-738-973-233/c
; Sequence 233, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Joseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 233
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (1)...(366)
; OTHER INFORMATION: n = A,T,C or G
US-09-738-973-233
```

```
Query Match          4.6%; Score 35; DB 10; Length 366;
Best Local Similarity 51.0%; Pred. No. 6;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

```
Oy 444 TATTCAAGATTGTTGTTGCTTGCCTTGAAATTTCTGATGAAGTAACATAACAT 503
    ||||| | ||||| || | ||||| || ||||| ||||| ||||| ||||| |||||
Db 263 TATTTAATATTGTTAATTTATTTATTCATGATAGAAATTTATGTAGAACAAACAAATACT 204
Oy 504 GACATCTTTTGAAGTACGACGATTCGCTAATGTTGTCAACCATATTGTCCTTTCAAT 563
    ||||| | ||||| || | ||||| || ||||| ||||| ||||| ||||| |||||
Db 203 TTATCCCACTTAAAGAAATATTAACATTTATGTACATATAATCTTTGTTTTTAAG 144
Oy 564 CTTCGATCCCAATTTTGGCGTTTATCTGATGATGATG 600
    ||||| | ||||| || | ||||| || ||||| ||||| ||||| ||||| |||||
Db 143 TTAGTGTAATTTTGTGTGATTAATCTTTTGTGTG 107
```

```
RESULT 31
US-09-939-980-236
; Sequence 236, Application US/09939980
; Patent No. US20020082234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratl, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,980
; FILING DATE: 27-Aug-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/936,165
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmil, Edward R
; REGISTRATION NUMBER: 38,891
; TELEPHONE/DOCKET NUMBER: P50549
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX: <unknown>
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; MOLECULE TYPE: Genomic DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:
US-09-939-980-236

```

Query Match	4.6%	Score 35;	DB 10;	Length 830;
Best Local Similarity	48.0%	Pred. No. 8.5'		
Matches 98; Conservative	0;	Mismatches 106;	Indels 0;	Gaps 0;

QY	144	TGATGCCATTCCTCTATTGGAACATCCGAAACATATGATTCACATGAATACCTGTCA	203
Db	471	TGAAGTATTTCACAGCATTTAGAACTATGAGAAAAAGACGACCAATTTTAAACATA	530
QY	204	TGCCGATPAATTCGAAGTCATGTTGATTTGAGATTTCAAAGGTAATTAAGCATGCG	263
Db	531	TGCGGAAAAATGATTGTTAAATTTCTATTGTTAAATACAAAAGGTGAATCAACGATTCC	590
QY	264	AAACATTGAGGCTCGAGGATTGAAACAAATGAAACGTCAAAGGTATCTTATGTCAAAG	323
Db	591	TTTAAATTAATAAAACCTTCACGTCGCAATGTGAGATTAAAGGTTTCGCTTTTACACAAAT	650
QY	324	TGAAGAGGTTATGTTAAAGCTCA	347
Db	651	TGAACAAAGTAAAGAAATTAACGTA	674

```

RESULT 32
US-09-938-842A-5067
: Sequence 5067, Application US/09398842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kireps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPT300-3
: CURRENT APPLICATION NUMBER: US/09/938, 842A
: CURRENT FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227, 866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264, 647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300, 111
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 5379
: SEQ ID NO 5067
: LENGTH: 2000
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
: US-09-938-842A-5067

```

Query Match	Best Local Similarity	4.6%	Score 35	DB 9	Length 2000
Matches 86	Conservative 0	Mismatches 85	Indels 0	Gaps 0	
QY 591	TGATGTATTGACCGCTATTTTCCAGACACCCCTAGCTAAGGAATGACCAAGTATTGGC				
Db 626	TGATGTTTTGTTATATATTTTCCCTATATACGTAACCTTCATTAAATCAATGGAATTGA				
QY 651	ACCAGCATTTAAACGTGTAATTGGAAAAAATTTAACCATAGACATCAITTTTCCACACTGT				
Db 686	AAGTATATTTGTCATTAGATAGTATAGTTAATTGAAATAAATAGCATTTATGCGTTAAATATAG				
QY 711	ACAATCTCTATTTCACGTACATAAATAAATTTTATTTATTTATTTTCC				
Db 746	CATGACGCTTTTCACTTAGAATCCATATCTATTTTTTTTTTTTTTTTGGC				

RESULT 33
US-09-070-927A-93/C
; Sequence 93, Application US/09070927A
; Patent No. US20020120116A1

```

GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
          Patrick J. Dillon
          Steven Barash
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 4178 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 93:
US-09-070-927A-93

```

[illegible]

RESULT 34
US-09-925-300-686
; Sequence 686, Application US/09925300

Db 701 AAGAGAGTAATCCAGATTGTCATTTGGATGTCAGAGTACCAATGGCAATGTTCCAAATA 760
Qy 386 AATATGATTAGCATACCAATTTGGGTGATCTTCA 419
Db 761 TATCTGATTTGTCATACAAAAGCCTTCCTGTGCA 794

RESULT 37
US-09-954-531-156
; Sequence 156, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand

; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 156
; LENGTH: 3143
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-156

Query Match 4.6%; Score 34.8; DB 9; Length 3143;
Best Local Similarity 46.1%; Pred. No. 17;
Matches 117; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 333 TATGTTAAGCTCATTTGTCATCGGTGTCACGATGATATGCTCTCGATGGAATATGA 392
Db 2846 TAATGCTTTGATTAATAGTTAGTTTGCGCTTCATGGAACCTCCTGTAACAAAGC 2905
Qy 393 TTTAGCATACAAATGGGTGATCTTCATCCACACATCATGTCTATTCGGATATTCAAGA 452
Db 2906 TTCAGGCTTATGCTATGTCATCTTATAGAGAAATGCAAACTATCAGTATTTTAAT 2965
Qy 453 TTTTGTGTTGCCCTGTCCCTTGAAATTTCTGATGAAGTATACATAACATGACATCTTT 512
Db 2966 ATTTGATATCTCATGATAGATAATTTATGTGAAGCAACAAATACCTTTTACCAC 3025
Qy 513 TGAAGTACGACAAATTCCTATATGTCACACATATGGTGGCTTTCAATCTTGATCC 572
Db 3026 TTAATAAGAGAAATTAACATTTATGTGACATAATCTTTTGTGTTTAAAGTATGTAT 3085
Qy 573 AATTTTGGCGTTT 586
Db 3086 ATTTTGTGTGATT 3099

RESULT 38
US-09-954-531-368
; Sequence 368, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand

; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133

; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 368
; LENGTH: 3143
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-368

Query Match 4.6%; Score 34.8; DB 9; Length 3143;
Best Local Similarity 46.1%; Pred. No. 17;
Matches 117; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 333 TATGTTAAGCTCATTTGTCATCGGTGTCACGATGATATGCTCTCGATGGAATATGA 392
Db 2846 TAATGCTTTGATTAATAGTTAGTTTGCGCTTCATGGAACCTCCTGTAACAAAGC 2905
Qy 393 TTTAGCATACAAATGGGTGATCTTCATCCACACATCATGTCTATTCGGATATTCAAGA 452
Db 2906 TTCAGGCTTATGCTATGTCATCTTATAGAGAAATGCAAACTATCAGTATTTTAAT 2965
Qy 453 TTTTGTGTTGCCCTGTCCCTTGAAATTTCTGATGAAGTATACATAACATGACATCTTT 512
Db 2966 ATTTGATATCTCATGATAGATAATTTATGTGAAGCAACAAATACCTTTTACCAC 3025
Qy 513 TGAAGTACGACAAATTCCTATATGTCACACATATGGTGGCTTTCAATCTTGATCC 572
Db 3026 TTAATAAGAGAAATTAACATTTATGTGACATAATCTTTTGTGTTTAAAGTATGTAT 3085
Qy 573 AATTTTGGCGTTT 586
Db 3086 ATTTTGTGTGATT 3099

RESULT 39
US-09-964-824A-557
; Sequence 557, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign

; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 557
; LENGTH: 3143
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-557

Query Match 4.6%; Score 34.8; DB 10; Length 3143;
Best Local Similarity 46.1%; Pred. No. 17;
Matches 117; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 333 TATGTTAAGCTCATTTGTCATCGGTGTCACGATGATATGCTCTCGATGGAATATGA 392
Db 2846 TAATGCTTTGATTAATAGTTAGTTTGCGCTTCATGGAACCTCCTGTAACAAAGC 2905
Qy 393 TTTAGCATACAAATGGGTGATCTTCATCCACACATCATGTCTATTCGGATATTCAAGA 452
Db 2906 TTCAGGCTTATGCTATGTCATCTTATAGAGAAATGCAAACTATCAGTATTTTAAT 2965
Qy 453 TTTTGTGTTGCCCTGTCCCTTGAAATTTCTGATGAAGTATACATAACATGACATCTTT 512
Db 2966 ATTTGATATCTCATGATAGATAATTTATGTGAAGCAACAAATACCTTTTACCAC 3025
Qy 513 TGAAGTACGACAAATTCCTATATGTCACACATATGGTGGCTTTCAATCTTGATCC 572
Db 3026 TTAATAAGAGAAATTAACATTTATGTGACATAATCTTTTGTGTTTAAAGTATGTAT 3085
Qy 573 AATTTTGGCGTTT 586
Db 3086 ATTTTGTGTGATT 3099

Db 2846 TAATGTGTTGAAVTAATAGTTAGTTTGTGCTTCATGGAACCTCCCTGTAACAAAGC 2905
Oy 393 TTACACATACAAATGGGTGATCTCTCACCACCACTCATGCAATTTGGATATTCAGA 452
Db 2906 TTCAGGTTATGCTATGTTCAATTCATAGAGAAATGCAACCTATCACTGATTTTAAT 2965
Oy 453 TTTTGTGTTGCTTGTCCCTTGAATTTTCGATGAGTAACATACAAATGACATCTTT 512
Db 2966 ATTGTATTTCTCATGATGAAATTTATGTAAGCAACAAATATCTTTACCCAC 3025
Oy 513 TGAATAGACAAATTCGCTAATGTTGTCAACCATATGCTGCTTTCAATTTGGATCC 572
Db 3026 TTAATAAGAGAAATTAACATTTATGTCACATAATCTTTGTTTAAAGTTAGTGAT 3085
Oy 573 AATTTTGGCGTTT 586
Db 3086 ATTTTGTGTGAT 3099

RESULT 40
US-09-880-107-3322
; Sequence 3322, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880.107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3322
; LENGTH: 3143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U20758
US-09-880-107-3322

Query Match 4.6%; Score 34.8; DB 10; Length 3143;
Best Local Similarity 46.1%; Pred. No. 17;
Matches 117; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Oy 333 TATTTTAAAGCTCATTTGTTGATCGGTTCACGATGATATCGTCTCGATGGAATATGA 392
Db 2846 TAATGTGTTGAAVTAATAGTTAGTTTGTGCTTCATGGAACCTCCCTGTAACAAAGC 2905
Oy 393 TTACACATACAAATGGGTGATCTCTCACCACCACTCATGCAATTTGGATATTCAGA 452
Db 2906 TTCAGGTTATGCTATGTTCAATTCATAGAGAAATGCAACCTATCACTGATTTTAAT 2965
Oy 453 TTTTGTGTTGCTTGTCCCTTGAATTTTCGATGAGTAACATACAAATGACATCTTT 512
Db 2966 ATTGTATTTCTCATGATGAAATTTATGTAAGCAACAAATATCTTTACCCAC 3025
Oy 513 TGAATAGACAAATTCGCTAATGTTGTCAACCATATGCTGCTTTCAATTTGGATCC 572
Db 3026 TTAATAAGAGAAATTAACATTTATGTCACATAATCTTTGTTTAAAGTTAGTGAT 3085
Oy 573 AATTTTGGCGTTT 586
Db 3086 ATTTTGTGTGAT 3099

RESULT 41
US-09-742-096-2

; Sequence 2, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773USODIV
; CURRENT APPLICATION NUMBER: US/09/742.096
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
; OTHER INFORMATION:
US-09-742-096-2

Query Match 4.5%; Score 34.6; DB 9; Length 5361;
Best Local Similarity 43.4%; Pred. No. 24;
Matches 160; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

Oy 91 GCTGATCCAAATTCATGATTAATTAATCAACGAAATCAACAAAGCTATGATGATGCC 150
Db 4582 GATGAAATAGTAGAAGTTGAAATGAACATGAGATATAGAAAGATGTAGAACAGAT 4641
Oy 151 ATTGCTGATTTGAACATCCGAAACATAGATCCAAATGAAGTACCTGATATGCCGAT 210
Db 4642 ATAGAAGAAATATAGAAAGATGAAAGTAAAGTTAGAAATAGATGAAGATATGATGAGAT 4701
Oy 211 AATTCGAACGTCATGTTGTTGTGATTTCAAGTGAATTTGCCATGCGAAACATT 270
Db 4702 ATAGCTGAAGACAAAGATGAGTTATGATTTAAATGATCAAAAAAGACAAACGCTTTGAA 4761
Oy 271 GAGGCTCGAGATTAACAAATGAACGTCAGGTGATGCTAATGTCAAAGTGAAGNG 330
Db 4762 AAGGTTAAAGCGAAAGAAAGAAATTAAGAAAAAGTTGAAGAGGTGATGATGCTTT 4821
Oy 331 GGTATTTGTTAAAGCTCATTTGTTGATCGGTGTTCAAGATGATATGCTCGATGGAATAT 390
Db 4822 AAAAAACAGTGAAGAGATATGAAATATGTTCAAAAAATGATTAAGAAAGTTGATAAA 4881
Oy 391 GATTTAGCATACAAATTTGGGTGATCTTCATCCACCACTCATGCTATTTGGATATTTCAA 450
Db 4882 GAAGTATCTAAAGCTTTAGAAATCAAAAAATGATGTTACTAATGTTTAAACAAATATCAA 4941
Oy 451 GATTTGTT 459
Db 4942 GATTTT 4950

RESULT 42
US-09-742-096-1
; Sequence 1, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773USODIV
; CURRENT APPLICATION NUMBER: US/09/742.096
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06

;; PRIOR APPLICATION NUMBER: PCT/FR96/00894
;; PRIOR FILING DATE: 1996-06-12
;; PRIOR APPLICATION NUMBER: FR 95/07007
;; PRIOR FILING DATE: 1995-06-13
;; NUMBER OF SEQ ID NOS: 29
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 1
;; LENGTH: 6152
;; TYPE: DNA
;; ORGANISM: P. falciparum
US-09-742-096-1

Query Match 4.5%; Score 34.6; DB 9; Length 6152;
Best Local Similarity 43.4%; Pred. No. 25;
Matches 160; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 91 GCGATCCATTCATGATGATAAATCCAGAGAAATCAACAAGCATATGATGATGCC 150
DB 4826 GATGAATTAAGTATGAGTGAATGAAGATGATGATGAGAGAGATGAGAGAGAT 4885
QY 151 ATTGCTGCTATTGACATCCGAACATAGATCAATGAAGTACCTGATGCCGAT 210
DB 4886 ATGAGAAGAAATATAGAGAAGATGAAGTGAAGATATGATGAAGATATGATGAAGAT 4945
QY 211 AAATTGCAACGTCATGTTGGTATTGATGATTCGAAGGTGAATGACCATGCGAACATT 270
DB 4946 ATAGGTGAAGACCAAGATGAAGTATGATTAATGATCCAAAAGAGAACCATGTA 5005
QY 271 GAGGCTCGAGATTTGAACAATGAACGTCAGTGATCTAATGCAAGGTGAAGAG 330
DB 5006 AAGGTTAAACGAAAAAGAAAAATTAAGAAAAAGTGAAGAGGTGTTAGTGCTTT 5065
QY 331 GGTATTGTTAAAGCTATTGTTGTCGTTCCAGCATATATCGCTCGATGGAATAT 390
DB 5066 AAAAAACGTAGACGAGATGAATGATGATGATGATGATGATGATGATGATGATGAT 5125
QY 391 GATTGACATCAAAATTTGGTGATCTTCATCCAAACCATGATGATGATGATGATGAT 450
DB 5126 GAAGTATCTAAAGCTTTAGATCAAAAAAATGATGATGATGATGATGATGATGATGAT 5185
QY 451 GATTGTTGTT 459
DB 5186 GATTGTTT 5194

RESULT 43
US-09-835-081-3
; Sequence 3, Application US/09835081
; Patent No. US20020151020A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Xianghe et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001224
; CURRENT APPLICATION NUMBER: US/09/835,081
; CURRENT FILING DATE: 2001-04-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 57130
; TYPE: DNA
; ORGANISM: Human
US-09-835-081-3

Query Match 4.5%; Score 34.6; DB 10; Length 57130;
Best Local Similarity 56.6%; Pred. No. 64;
Matches 64; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 633 AATGACCAAGTATTGACACAGATTTAAACGATTTGGAAGAAAAATTAACCAATAGA 692
DB 8662 AAGGTTGATGATGTTAGCATCTTTTAACAAGAAATTAATTAAGTATGTA 8721

QY 693 CATCATTTTTCCACGTGACATCTCTATTTCAGTGAATTAATAAATTTT 745
DB 8722 CATGTTTATTAGCTATGCTATTAATGACACTGATGACGATGAT 8774

RESULT 44
US-09-974-300-6524
; Sequence 6524, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6524
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(624)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-6524

Query Match 4.5%; Score 34.2; DB 10; Length 624;
Best Local Similarity 48.7%; Pred. No. 12;
Matches 93; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 23 AAAAAACATATCTTCAAAATGATGAATTTTGTGATGTCGCGGATTTGTCG 82
DB 36 AATCAACAGTGTAAATAATGATGATGATGATGATGATGATGATGATGATGATGAT 95
QY 83 CCGTTTCGCTGATCCATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 142
DB 96 TTTTGGGCTGACATTAACCGATGAGAGCGATCAAGCGCGATTTGATTTGTTT 155
QY 143 ATGATGCAATTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 202
DB 156 ATGATGCAATGTTTTTTTGAAGGCTGATTAAGATTAAGATTAAGATGATGATGAT 215
QY 203 ATGCCGATTA 213
DB 216 CAGCTTACAA 226

RESULT 45
US-09-765-272-95
; Sequence 95, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2


```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4696
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4696
```

```
Query Match 4.5%; Score 34; DB 10; Length 618;
Best Local Similarity 45.0%; Pred. No. 14;
Matches 127; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 92 CTGATCCAAATTCATGATGATTAATAATCACGAGAAATCAACAAGCTATTGATGATGCCA 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 CTGAACAATTAATAATAGAGGTGTAAGCTTTCAGAGAAATGAATCACTAGATGATCAAA 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 TTGCTGCTATTGAACAATCCGAAACATAGATCCATGAATGAATGATCATGATCCGATA 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 ATATTTCATTTGACGACACACTTTCATTCAGCTCCGAACATATCAATTTACAATCTTA 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 AATTCGAAGCTCATGTTGATTTGTGATTTCAAGGTGAATTAAGCCATGCGAAACATTTG 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 AAGCAGCCAAAAGAGTAGATGCTTAAAGTTGTAATGAACAGTAAGTATA 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 AGGCTCAGAGATTGAACAATGAACGTCAAGGTGATGCTAATGTCAAAAGGTGAAGAG 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 AATGACAAAGTATTAAGATGTTAAGCTTACAGATGTAGAAGTCTGATGAACACAAAAG 520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 GTATTGTTAAAGCTCATTTGTTGATCGTGTTCAGATGATA 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 GTAAAGATTAACAATTAACATTAATTAATTGATGATTAACA 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 49
US-09-933-999A-2
; Sequence 2, Application US/09933999A
; Publication No. US20030022178A1
; GENERAL INFORMATION:
; APPLICANT: U.C. Regents
; TITLE OF INVENTION: Identification of Sortase Gene
; FILE REFERENCE: 510015-249
; CURRENT APPLICATION NUMBER: US/09/933,999A
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-933-999A-2
```

```
Query Match 4.5%; Score 34; DB 9; Length 621;
Best Local Similarity 45.0%; Pred. No. 14;
Matches 127; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 92 CTGATCCAAATTCATGATGATTAATAATCACGAGAAATCAACAAGCTATTGATGATGCCA 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 CTGAACAATTAATAATAGAGGTGTAAGCTTTCAGAGAAATGAATCACTAGATGATCAAA 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 TTGCTGCTATTGAACAATCCGAAACATAGATCCAAATGAAGTACCGATGATCCGATA 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 ATATTTCATTTGACGACACACTTTCATTCAGCTCCGAACATATCAATTTACAATCTTA 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 AATTCGAAGCTCATGTTGATTTGTGATTTCAAGGTGAATTAAGCCATGCGAAACATTTG 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 AAGCAGCCAAAAGAGTAGATGCTTAAAGTTGTAATGAACAGTAAGTATA 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 AGGCTCAGAGATTGAACAATGAACGTCAAGGTGATGCTAATGTCAAAAGGTGAAGAG 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 AATGACAAAGTATTAAGATGTTAAGCTTACAGATGTAGAAGTCTGATGAACACAAAAG 520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 GTATTGTTAAAGCTCATTTGTTGATCGTGTTCAGATGATA 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 GTAAAGATTAACAATTAACATTAATTAATTGATGATTAACA 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 50
US-09-933-999A-39/C
; Sequence 39, Application US/09933999A
; Publication No. US20030022178A1
; GENERAL INFORMATION:
; APPLICANT: U.C. Regents
; TITLE OF INVENTION: Identification of Sortase Gene
; FILE REFERENCE: 510015-249
; CURRENT APPLICATION NUMBER: US/09/933,999A
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-933-999A-39
```

```
Query Match 4.5%; Score 34; DB 9; Length 621;
Best Local Similarity 45.0%; Pred. No. 14;
Matches 127; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 92 CTGATCCAAATTCATGATGATTAATAATCACGAGAAATCAACAAGCTATTGATGATGCCA 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 CTGAACAATTAATAATAGAGGTGTAAGCTTTCAGAGAAATGAATCACTAGATGATCAAA 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 TTGCTGCTATTGAACAATCCGAAACATAGATCCATGAATGAATGATCATGATCCGATA 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 ATATTTCATTTGACGACACACTTTCATTCAGCTCCGAACATATCAATTTACAATCTTA 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 212 AATTCGAAGCTCATGTTGATTTGTGATTTCAAGGTGAATTAAGCCATGCGAAACATTTG 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 221 AAGCAGCCAAAAGAGTAGATGCTTAAAGTTGTAATGAACAGTAAGTATA 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 AGGCTCAGAGATTGAACAATGAACGTCAAGGTGATGCTAATGTCAAAAGGTGAAGAG 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 AATGACAAAGTATTAAGATGTTAAGCTTACAGATGTAGAAGTCTGATGAACACAAAAG 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 GTATTGTTAAAGCTCATTTGTTGATCGTGTTCAGATGATA 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 GTAAAGATTAACAATTAACATTAATTAATTGATGATTAACA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: February 20, 2003, 22:17:02
Job time : 696 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2003, 20:12:57 ; Search time 1472 Seconds
(without alignments)
8372.803 Million cell updates/sec

Title: US-10-024-955-6
Perfect score: 761
Sequence: 1 GATCTATATCATACATACAT.....ATTATATTTATTTCTCC 761

Scoring table: IDENTITY_NUC
Gapex 10.0, Gapext 1.0

Searched: 16154066 segs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estmu: *
5: em_estcov: *
6: em_estcpl: *
7: em_estcro: *
8: em_hlc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hlc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rnd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61.2	8.0	506	14	B0834778 B0834778 Po.ad_03C
2	61.2	8.0	601	14	B0835021 B0835021 Po.ad_06C
3	61	8.0	598	14	B0835006 B0835006 Po.ad_06B
4	52	6.8	541	13	B0278674 AS_t92_65
5	52	6.8	556	12	B0169380 AS_t92_24
6	46.8	6.1	700	17	BH207437 Sml-43p13

C	7	46.8	6.1	779	17	BH200189	BH200189 Sml-56L14
C	9	46.8	6.1	856	17	BH178274	BH178274 011_F_07-
C	8	46.8	6.1	856	17	CNS07KML	AL615231 T7 end of
C	10	46.6	6.1	900	17	A2549980	A2549980 ENTDP94TF
C	11	45.6	6.0	905	17	A2550256	A2550256 ENTVE58TR
C	12	45.4	6.0	912	17	A2551092	A2551092 ENTFC22TF
C	13	45	5.9	816	17	BH178682	BH178682 012_F_13-
C	14	45	5.9	816	17	CNS07KXX	AL615639 T3 end of
C	15	44.4	5.8	916	17	A2671886	A2671886 ENTRH44TR
C	16	44.4	5.8	1101	17	CNS00E14	AL069257 Drosophila
C	17	44.2	5.8	292	17	A0076852	A0076852 CTR-HSP-2
C	18	43.8	5.8	341	13	BH165203	BH165203 EST567726
C	19	43.8	5.8	456	13	BH160594	BH160594 EST563117
C	20	43.8	5.8	890	17	A2530768	A2530768 ENTRH54TF
C	21	43.6	5.7	525	17	BH184460	BH184460 025_B_19-
C	22	43.6	5.7	525	17	CNS07PEA	AL621412 T7 end of
C	23	43.4	5.7	843	17	A2551618	A2551618 ENTRV54TR
C	24	43.4	5.7	908	17	A2548467	A2548467 ENTRK30TR
C	25	43.4	5.7	931	17	BH160272	BH160272 ENTQV49TR
C	26	43.2	5.7	581	13	B1747040	B1747040 rm5b05.Y
C	27	42.8	5.6	563	13	B1371247	B1371247 B1371247
C	28	42.6	5.6	483	13	BJ369346	BJ369346 BJ369346
C	29	42.6	5.6	484	13	BJ370579	BJ370579 BJ370579
C	30	42.6	5.6	489	13	BJ391279	BJ391279 BJ391279
C	31	42.6	5.6	506	13	BJ365024	BJ365024 BJ365024
C	32	42.6	5.6	516	13	BJ388204	BJ388204 BJ388204
C	33	42.6	5.6	519	13	BJ368463	BJ368463 BJ368463
C	34	42.6	5.6	525	13	B1594832	B1594832 AS_t92_44
C	35	42.6	5.6	533	13	BJ367414	BJ367414 BJ367414
C	36	42.6	5.6	537	13	BJ334321	BJ334321 BJ334321
C	37	42.6	5.6	538	13	BJ330406	BJ330406 BJ330406
C	38	42.6	5.6	540	13	BJ365955	BJ365955 BJ365955
C	39	42.6	5.6	551	13	BJ367714	BJ367714 B1367714
C	40	42.6	5.6	559	13	BJ339394	BJ339394 BJ339394
C	41	42.6	5.6	559	13	BJ365574	BJ365574 BJ365574
C	42	42.6	5.6	567	13	BJ328350	BJ328350 BJ328350
C	43	42.6	5.6	574	13	BJ370337	BJ370337 BJ370337
C	44	42.6	5.6	582	13	BJ335809	BJ335809 BJ335809
C	45	42.6	5.6	582	13	BJ366050	BJ366050 BJ366050

ALIGNMENTS

RESULT 1
B0834778
LOCUS
DEFINITION
B0834778 506 bp mRNA linear EST 08-AUG-2002
Po.ad_03C12.7EXPI Psoroptes ovis mixed Psoroptes ovis cDNA clone
Po.ad_03C12.5' similar to Q9UG2 PUTATIVE ALLERGEN PRECURSOR.
Lepidoglyphus destructor, mRNA sequence.

ACCESSION
B0834778
VERSION
B0834778.1 GI:22139092

KEYWORDS
SOURCE
ORGANISM

Psoroptes ovis
sheep scab mite.
Psoroptes ovis
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Sarcopitidae; Psoroptidae;
Psoroptes.
1 (bases 1 to 506)
Kenyon, F., Welsh, M., Parkinson, J., Whitton, C., Blaxter, M. and Knox
D.P.
Expressed sequence tag survey of gene expression in the scab mite
Psoroptes ovis - allergens, proteinases and free radical scavengers
Unpublished (2002)
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JF, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

The library was prepared for Dr David Knox, Moredun Research

psoroptes.
1 (bases 1 to 598)
Kenyon, F., Welsh, M., Parkinson, J., Whilton, C., Blaxter, M. and Knox, D.P.

OY	480	TTCGATG 487
Dn	154	TACTGTG 147
<hr/>		
RESULT 5	Bf169380/c	
LOCUS	Bf169380	
DEFINITION	A5_tg_24F05.SKPL Ascaris suum adult male testis germinal zone from Alan Scott	
ACCESSION	Bf169380	
VERSION	Bf169380..1 GI:11053992	
KEYWORDS	EST.	
SOURCE	pig roundworm.	
ORGANISM	Ascaris suum Eukaryota; Metazoa; Nematoda; Chromadorea; Ascariida; Ascarididae; Ascariidae; Ascaris. 1 (bases 1 to 556) Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Gulliano,D., Hall, N., Quayle,M. and Barrell,B. Edinburgh University/Sanger Centre Nematode EST Project Unpublished (2000)	
TITLE	Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK. Tel.: +44 131 650 6760 Fax: +44 131 670 5450 Email: mark.blaxter@ed.ac.uk	
JOURNAL	The library was prepared by Michelle Lizotte-Maniewski for Alan Scott. Johns Hopkins University Medical School, Baltimore. Sequencing was performed by Claire Whitton ICAPB, Edinburgh The sequence contained a Polya tail (trimmed) PCR Primers FORWARD: T3 BACKWARD: T7PL Plate: 24 row: F column: 05 Seq primer: SKPL High quality sequence stop: 543. Location/Qualifiers	
FEATURES	location/Qualifiers 1..556 /organism="Ascaris suum" /db_xref="taxon:6253" /clone="AS-tg_24F05" /clone_lib="Ascaris suum adult male testis germinal zone from Alan Scott" /sex="Male" /dev_stage="Adult" /note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI; Library was made from dissected testis germinal zone from adult male Ascaris suum collected from abattoirs. Constructed by Michelle Lizotte-Maniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD."	
BASE COUNT	202 a 156 c 79 g 119 t	
ORIGIN		
Query Match	6.8%; Score 52; DB 12; Length 556;	
Best Local Similarity	48.1%; Pred. No. 0.019;	
Matches 148:	Conservative 0; Mismatches 160; Indels 0; Gaps 0;	
OY	180	AGATCCATTGAAGAAGTGGATCATCGCAGATAAATTCACAAGCTCATTTGTAATGTGSA 239
Dn	469	AGTGCGATTAAAGTGTGC GTTGTCAAATTTGTAAAATTTGCTGTGAAGTGTCTGT 410
OY	240	TTTTCAAGTGAATTAGCATGCAAAACATTGAGGCTCGAGATTGAAAAAATGAACG 299
Dn	409	TGTTGAATTTGCTGTGTGATTTGAATATGTCGTCTTGTGAAGTGTGTTGAAGTTACTGC 350
OY	300	TCAACGTATGCTAAATGTCAAAGGTGAAGAGGATTTGTTAAAGCTCATTTGTGATCGG 359
Dn	349	TGCTGAAGTTCGTGTGCTGAAGTTCGAGGTTAATTTGATGTGAATGCTGAAGTTGTAGT 290

QY	360	TGTTTCACGATCATATACCTCGATGAGTAATATTATTTACCATACAAATATGGGTGATCTTCA	419
Db	289	TAGAGCTACTGAGGGTTGTTGTTGAGGTTTGAGGTTCTTCAAGTTAGTAGAGTTGT	230
QY	420	TCCAACACATCATATGTCATTTGGGATATTCAAGATTTTGTGCTTGTCCCTTGAAT	479
Db	229	TCAGGAGCTATGTTGAGAGTTACTGAAACAGCTGAGGTTATTATTGAGGTTGCTGTAAT	170
QY	480	TTTCGATG 487	
Db	169	TACTGTTG 162	
RESULT 6			
EH207437/c			
LOCUS	BH207437	700 bp	DNA linear GSS 24-OCT-2001
DEFINITION	Sml-43Pl3.TF Sml Schistosoma mansoni genomic clone Sml-43Pl3, DNA		
ACCESSION	BH207437		
VERSION	BH207437.1	GI:16384444	
KEYWORDS	GSS.		
ORGANISM	Schistosoma mansoni.		
SOURCE	Schistosoma mansoni		
REFERENCE	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidae; Schistosomatoidae; Schistosomatidae; Schistosoma.		
AUTHORS	Shetty,J., Simpson,A., Malek,J., Koo,H., Loverde,P.T. and El-Sayed		
TITLE	'N.M.		
JOURNAL	Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sml BAC library for gene discovery and map construction		
COMMENT	Unpublished (2001) Other GSSs: Sml-43Pl3.TF Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: nelsayed@tigr.org		
FEATURES	lo.edu). Seq primer: M13 For Class: BAC ends.		
source	Location/Qualifiers 1..700 /organism="Schistosoma mansoni" /strain="Puerto Rico" /db_xref="taxon:6183" /clone="Sml-43Pl3" /clone_id="Sml" /note="Vector: pBelOBAC11; Site.1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CDRP, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."		
BASE COUNT	233 a 169 c 90 g 208 t		
ORIGIN			
Query Match	6.1%; Score 46.8; DB 17; Length 700;		
Best Local Similarity	45.8%; Pred. No. 0.38; 192; Indels 0; Gaps 0;		
Matches 162; Conservative 0; Mismatches 192; Indels 0; Gaps 0;			
QY	40	AAAATGATGAATTTTCTGATTGCTCCGCTGGCATTTTGCACGCTTTCGGCTGATCCA	99
Db	408	AATATAATGATGAAATATTTCTATGCTCTGTTGCTGCTGCTGATGATCTATGCT	349
QY	100	ATTCACTATGATTAATACCGAACAATTCACAAAGCTATTATGATGATGCCATTGCTCT	159

Db 348 GCTGCTGATGCTAATGCTGCTGCTAATGCTGCTGCTGCTGCTGCTGCTGCT 289
Oy 160 ATGACATCCGAACAATAGATCCATGAAAGTACTGATCATCCGATTAATCCGA 219
Db 288 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 229
Oy 220 CGTCATGTTGGTATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 279
Db 228 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 169
Oy 280 GGATGAAACAATGAAACGTCAGATGATGATGATGATGATGATGATGATGATGAT 339
Db 168 TTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 109
Oy 340 AAAGCTATTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 393
Db 108 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 55

RESULT 7

LOCUS BH200189 779 bp DNA linear GSS 24-OCT-2001
DEFINITION Sml-56L14.TR Sml Schistosoma mansoni genomic clone Sml-56L14, DNA
sequence.
ACCESSION BH200189
VERSION BH200189.1 GI:16370307
KEYWORDS GSS.
SOURCE Schistosoma mansoni.
ORGANISM Schistosoma mansoni.
REFERENCE Eukaryota: Metazoa: Platyhelminthes: Trematoda: Digenea:
Strigoida: Schistosomatidae: Schistosomatidae: Schistosoma.
1 (bases 1 to 779)
AUTHORS Shetty,J., Simpson,A., Malek,J., Koo,H., Loverde,P.T. and El-Sayed
,N.M.

TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)
JOURNAL Sml BAC library for gene discovery and map construction
COMMENT Unpublished (2001)
Other GSSs: Sml-56L14.TF

CONTACT: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
lo.edu).

Seq primer: M13 Rev
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..779

/organism="Schistosoma mansoni"
/strain="Puerto Rico"
/db_xref="taxon:6183"
/clone="Sml-56L14"
/clone_1id="Sml"
/note="Vector: pBelobAC11; Site_1: Hin dIII; Constructed
in the laboratory of Dr. Denis Le Paslier at the Fondation
Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma
mansoni agarose embedded DNA was partially digested with
Hin dIII. High molecular weight fragments were ligated in
pBelobAC11 digested with Hin dIII. The average insert size
is 100 kb. Total clone coverage: approx. 7.95 x the
haploid genome. Further information can be found in Le
Paslier et al. (2000) Construction and characterization of
a Schistosoma mansoni bacterial artificial chromosome
library. Genomics 65: 87-94."

BASE COUNT 256 a 184 c 103 g 236 t
ORIGIN

Query Match 6.1%; Score 46.8; DB 17; Length 779;
Best Local Similarity 45.8%; Pred. No. 0.38;
Matches 162; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

Oy 40 AAATGATGAATTTTGTGATGCTCCGCGCATTTGCGCGTTCGATCCGA 99
Db 403 AATATATGATGATATATTTCTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
Oy 100 ATTCACATGATAAATACCGAAGAAATCAACAAAGCTATTGATGATGATGATGAT 159
Db 343 GCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 284
Oy 160 ATGACATCCGAACAATAGATCCATGAAAGTACTGATCATCCGATTAATCCGA 219
Db 288 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 224
Oy 220 CGTCATGTTGGTATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 279
Db 228 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 164
Oy 280 GGATGAAACAATGAAACGTCAGATGATGATGATGATGATGATGATGATGATGAT 339
Db 168 TTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 104
Oy 340 AAAGCTATTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 393
Db 103 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 50

RESULT 8

LOCUS BH178274 856 bp DNA linear GSS 19-OCT-2001
DEFINITION 011_F_07-21 SmbAC1 Schistosoma mansoni genomic clone 011F07 5', DNA
sequence.
ACCESSION BH178274
VERSION BH178274.1 GI:16277893
KEYWORDS GSS.
SOURCE Schistosoma mansoni.
ORGANISM Schistosoma mansoni.
REFERENCE Eukaryota: Metazoa: Platyhelminthes: Trematoda: Digenea:
Strigoida: Schistosomatidae: Schistosomatidae: Schistosoma.
1 (bases 1 to 856)
AUTHORS Le Paslier,M.-C., Pierre,R.J., Merlin,F., Hirrl,H., Wu,W., Williams
,D.L., Johnston,D., Loverde,P.T. and Le Paslier,D.

TITLE Construction and characterization of a Schistosoma mansoni
JOURNAL bacterial artificial chromosome library
MEDLINE Genomics 65 (2), 87-94 (2000)
COMMENT Other GSSs: 011_F_07-rev
Contact: Pierre RJ
INSERM U 167
Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=D60AA011CC04CPI
Plate: 011 row: F column: 07
Seq primer: M13 -21 primer
Class: BAC ends
High quality sequence stop: 856.

FEATURES
source Location/Qualifiers
1..856

/organism="Schistosoma mansoni"
/strain="Puerto Rican"
/db_xref="taxon:6183"
/clone="011F07"
/clone_1id="SmbAC1"
/sex="mixed"
/dev_stage="cercariae"
/lab_host="Biomphalaria glabrata"
/note="Vector: pBelobAC 11; Site_1: Hind III; Partially
Hind III digested and size-selected S. mansoni cercarial
DNA was ligated into Hind III digested pBelobAC 11 vector
and used to transform E. coli DH10B. The complete library
contains 23808 clones from 4 independent

77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).

BASE COUNT 376 a 63 c 235 g 226 t

Email: bjoftuse@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 100
High quality sequence stop: 872.
Location/Qualifiers
1. 905

Query Match 6.1%; Score 46.6; DB 17; Length 900;

Best Local Similarity 44.8%; Pred. No. 0.43;

FEATURES
SOURCE
1. 905
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS1. Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

Matches 221: Conservative 0; Mismatches 269; Indels 3; Gaps 1;

Query Match 6.0%; Score 45.6; DB 17; Length 905;
Best Local Similarity 45.0%; Pred. No. 0.76;
Matches 171: Conservative 0; Mismatches 209; Indels 0; Gaps 0;

5 TATATCAATTAACATCCAAAAACATATCTTACAAATGATGAAATTTTGTGATG 64
122 TTTTATACACATCTCAAAAAGCATATTTTATGATTTAATGACACATTTTATTTG 181
65 CTGCGGTGACATTTTGTGCGGCTGATCCATTCATGATGATTAATCACCAG 124
182 TTTTACTGCTTTGTTCTGCTGCGGCGAGAGAGAGAAAGTTTGAATTAAGAAG 241
125 AATTAACAACATATGATGATGATGATGATGATGATGATGATGATGATGATG 184
242 ACTTCAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATG 301
185 CAATGAAGTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 244
302 TTGACTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
245 AAGTGAATTAACATCCAAAAACATATCTTACAAATGATGAAATTTTGTGATG 304
362 AGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
305 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 364
422 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481
365 AGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
482 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 541
422 CAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481
542 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
482 CTGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
602 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 614

RESULT 11
LOCUS A2550256 905 bp DNA linear GSS 14-NOV-2000
DEFINITION ENTEVE58TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION A2550256
VERSION A2550256.1 GI:11175557
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 905)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).

BASE COUNT 203 a 260 c 46 g 396 t

Query Match 6.0%; Score 45.6; DB 17; Length 905;
Best Local Similarity 45.0%; Pred. No. 0.76;
Matches 171: Conservative 0; Mismatches 209; Indels 0; Gaps 0;

111 TAAATCACCAGAAATTAACAAAGCTATTGATGATGATGATGATGATGATGATG 170
649 TGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 590
171 CGAAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 230
589 TGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 530
231 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 290
529 TGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 470
291 AATGAACCTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 350
469 TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 410
351 GTTGAATCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 410
409 TGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 350
411 TGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 470
349 TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 290
471 CCTGAATTTCTGATGAAG 490
289 AGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 270

RESULT 12
LOCUS A2551092 912 bp DNA linear GSS 14-NOV-2000
DEFINITION ENTFJ22TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION A2551092
VERSION A2551092.1 GI:11176393
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 912)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

	AUTHORS	1 (bases 1 to 912)
	TITLE	Lofthus B., Van Aken S. and Fraser C.
JOURNAL	Determination of clone end sequences from Entamoeba histolytica	
COMMENT	HMI:IMSS sheared DNA library Unpublished (2000) Contact: Brendan J Lofthus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjl@lofususelgr.org Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library	
FEATURES	Seq primer: M13-Forward Class: shotgun High quality sequence start: 17 High quality sequence stop: 861. Location/Qualifiers 1..912 /organism="Entamoeba histolytica" /strain="HMI:IMSS" /db_xref="taxon:5759" /clone_lib="Entamoeba histolytica Sheared DNA" /note="Vector: pHOSt1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O., and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Bareilly, Oxford University Press, 1999)."	
BASE COUNT	423 a 34 c 264 g 191 t	
ORIGIN		
Query Match	6.0%; Score 45.4; DB 17; Length 912;	
Best Local Similarity	45.4%; Pred. No. 0.85;	
Matches 163; Conservative	0; Mismatches 196; Indels 0; Gaps 0;	
OY	102 TCACATGATAAATCCACGAAGAATCAACAAGCTATTGATGCCATTGCTGTAT	161
Db	533 TGAAGATGATGATGAAGAAGAAGAAAGATGATGAAGAAGATGATGAAGAAGATGATGA	592
OY	162 TGAACATCCGAACAAATAGATCCATGAATACCCTGTCATGGCGATAAATTGCAGC	221
Db	593 TGAAAGAACAGAGATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGAT	652
OY	222 TCATGTTGTAATTCGATTTCCAAGTGTAATTAGCATGCCAATTCAGAGCTCGAG	281
Db	653 AGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGATGAAGA	712
OY	282 ATTGAACAATGAAGAGCTCAAAGTGATGCTATGTCAAAGSTGAAGAGGATTTGTTAA	341
Db	713 TGATGATGAAGAAGAAGATGATGATGATGAAGAAGAAGATGATGATGATGAAGA	772
OY	342 AGCTATTTGTTGATCGGTTCCACAGATGATTCCTCTCGATGGAATATGATTTAGCATTA	401
Db	773 AGAAGATGATGATGAAGAAGAAGATGATGAAGAAGAAGATGATGATGAAGA	832
OY	402 CAATTTGGTGATCTTCACCAACACATCATCTTCCTCGATATTCAAGATTTTGGTG	460
Db	833 AGAAGATGATGATGAAGAAGACGACGAATATGTAATTTTAGATGATGAAGAAGATGATG	891
RESULT 13	BH178682/c	BH178682 816 bp DNA linear GSS 19-OCT-2001
LOCUS	012_F_13-rev SmBAC1 Schistosoma mansoni genomic clone 012F13 5'	
DEFINITION		

	ACCESSION	DNA sequence.
	BH178682	
	BH178682.1	GI:16278732
	VERSION	
	KEYWORDS	GSS:
	SOURCE	Schistosoma mansoni.
	ORGANISM	Schistosoma mansoni Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida; Schistosomatidae; Schistosomidae; Schistosoma.
	REFERENCE	Le Paslier,M.-C., Pierce,R.J., Merlin,F., Hiral,H., Wu,W., Williams, J.D., Johnston,D., Loverie,P.T. and Le Paslier,D. Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library Genomics 65 (2), 87-94 (2000)
	JOURNAL	
	MEDLINE	20247247
	COMMENT	Other GSSs: 012.F.13-21 Contact: Pierre RJ INSERM U 167 Institut Pasteur de Lille 1 rue du Professeur A. Calmette, 59019-Lille, France Tel: (33) (0)3 20877783 Fax: (33) (0)3 20877888 Email: Raymond.pierce@pasteur-lille.fr CNS sequencing ID=DSMA012CC07BP1 Plate:012 row: F column: 13 Seq primer: M13 reverse primer Class: BAC ends High quality sequence stop: 816. Location/Qualifiers
FEATURES		
source		1..816 /organism="Schistosoma mansoni" /strain="Puerto-Rican" /db_xref="taxon:6183" /clone="012F13" /clone_1id="SmBAC1" /sex="mixed" /dev_stage="cercariae" /lab_host="Blomphalaria glabrata" /note="Vector: pReloBAC 11; Site_1: Hind III; Partially Hind III digested and size-selected S. mansoni cercarial DNA was ligated into Hind III digested pReloBAC 11 vector and used to transform E. coli DH10B. The complete library contains 23808 clones from 4 independent sizeing-ligation-transformations. Average insert size ranges from 70-127 kb and genome coverage is 7.9-fold."
BASE COUNT	264 a 190 c 116 g 244 t	others 2
ORIGIN		
Query Match	5.9%; Score 45;	DB 17; Length 816;
Best Local Similarity	47.3%;	Pred. No. 1.1;
Matches 169;	Conservative 0;	Mismatches 185; Indels 3; Gaps 1,
OY	40 AAAATGATGAATTTTGTGTCATCTGCTCGCGGCATTGTGCGCGTTGGCGTGATCCA	99
bD	384 AATAAATGAGAAATVCTTCTATCTCTGCTGTTGGTGGCTGCTCATCTTAATGCT	325
OY	100 ATTCACCTGATTAATCAACGGAAMAATCAACAAGCATATGATGATGCCATTGCTGCT	159
bD	324 GCTGCTGATGCTAAAGCTGCTGCTAATGCTGATGCTGCTGCTGCTGATGCTGCCGCTGCT	265
OY	160 ATTGAACAATCCGAACAATAGATCAATGAAGAAGTACTGATCATGCCGATAAATTCGAA	219
bD	264 TATGATGATGATGATCCCTGCTGCTCTAATGCTGCTGCTGCTGCTGCTGCTGATGATGAT	205
OY	220 CGTCATGTTGGTATGTCGATTTCAAAGTGAATTAGCCATCCGAACATTAGGCTCGA	279
bD	204 GATGATGATGATGATGATGAAGATATGATGATGAAGTTGAAGAAAGGATATATAA	145
OY	280 GGATTTGAACAACAAA---TGAACGTCACAGCTGCTGCTAATGTCGAAGGTCGAAGGATATT	336
bD	144 GATTTTGGACATGATGCTGTTTGTGACGATGACGATGATGATGATGATGATGAACATTAAG	85
OY	337 GTTAAAGCTCATTTTGTGATCGGTGTTACGATGATATGCTGTCGATGATGATATGAT	393


```

Db      84 GGATATGATGACTTATGTGACAAATATGATATGATGTTGACAGAGCGAAGCGGAT 28

RESULT 14
LOCUS   A2671886/c
DEFINITION
  T3 end of clone 012CC07 of library SmbAC1 from strain Puerto-Rican
  of Schistosoma mansoni, genomic survey sequence.
ACCESSION
  AL615639
VERSION
  AL615639.1 GI:16028863
KEYWORDS
  GSS.
SOURCE
  Schistosoma mansoni.
  Schistosoma mansoni.
  Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
  Strigedida; Schistosomatidae; Schistosomatidae; Schistosoma.
REFERENCE
  1 (bases 1 to 816)
  Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
  Williams,D.L., Johnston,D., Loverde,P.T. and Le Paslier,D.
  Construction and characterization of a Schistosoma mansoni
  bacterial artificial chromosome library
  Genomics 65 (2), 87-94 (2000)
JOURNAL
  MEDLINE
  20247247
PUBMED
  10783255
REFERENCE
  2 (bases 1 to 816)
AUTHORS
  Genoscope.
TITLE
  Direct Submission
  Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
  - Web : www.genoscope.cns.fr)
COMMENT
  Partially Hind III digested and size-selected S. mansoni cercarial
  DNA was ligated into Hind III digested pBlotBAC 11 vector and used
  to transform E. coli DH10B. The complete library contains 23808
  clones from 4 independent sizing-ligation-transformations. Average
  insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
  Location/Qualifiers
    source
      1. 816
        /organism="Schistosoma mansoni"
        /strain="Puerto-Rican"
        /db_xref="taxon:6183"
        /clone="012CC07"
        /clone_11b="SmbAC1"
        /note="end : 73"
        /note="end : 73"
BASE COUNT
  264 a 190 c 116 g 244 t 2 others
ORIGIN
  Query Match 5.9%; Score 45; DB 17; Length 816;
  Best Local Similarity 47.3%; Pred. No. 1.1;
  Matches 169; Conservative 0; Mismatches 185; Indels 3; Gaps 1;

QY 40 AAAAAAGTGAATTTTGTGATGTCGCCGCGATTTGTCGCCGTTCCGCTGATCCA 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 AATATAATGATGAATATGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 325
QY 100 ATTCACTATGATAAATCAACGAGAAATCAACAAAGCTATTGATGATGCTGCT 159
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 GCTGCTGATGCTAATGCTGCTGCTAATGCTGATGCTGCTGCTGCTGCTGCTGCT 265
QY 160 ATTGAACAATCCGAACAATAGATCCATGAAGTACTGATGATGCTGATGATGCTGAA 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 TATGATGATGATGATGCTGCTGCTGCTAATGCTGCTGCTGCTGCTGCTGATGAT 205
QY 220 CGTCATGTTGATGATGATGATTTCAAGGTGAATTAAGCATGCGAAACCTTTAGCTCA 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 GATGATGATGATGATGATGAAGATGATGATGATGAAGATGAAGATGAAGATGAAGA 145
QY 280 GGATGGAACAA--TGAAACGTCAGGTGATGCTAATGTCAAAGGTGAAGAGGTATT 336
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 GATTTTGACATGATGTTGTTTGTGACGATGACGATGATGATGATGATGATGAAGTAAG 85
QY 337 GTTAAAGCTATTTGTTGATCGGTGTTCAAGATGATATGCTGATGATGATGATGAT 393
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 GGATATGATGACTTATGTGACAAATATGATATGATGTTGACAGAGCGAAGCGGAT 28

```

```

RESULT 15
LOCUS   A2671886/c
DEFINITION
  ENTLR44TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
  genomic, DNA sequence.
ACCESSION
  A2671886
VERSION
  A2671886.1 GI:11809032
KEYWORDS
  GSS.
SOURCE
  Entamoeba histolytica.
  Entamoeba histolytica.
  Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
  1 (bases 1 to 916)
  Loftus,B., Van Aken,S. and Fraser,C.
  Determination of clone end sequences from Entamoeba histolytica
  HMI:IMSS sheared DNA library
  Unpublished (2000)
JOURNAL
  COMMENT
  Contact: Brendan J Loftus
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0208
  Fax: 301 838 3543
  Email: bjoftus@eigf.org
  Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
  DNA library
  Seq primer: M13-Reverse
  Class: Shotgun
  High quality sequence start: 92
  High quality sequence stop: 672.
  Location/Qualifiers
    source
      1. 916
        /organism="Entamoeba histolytica"
        /strain="HMI:IMSS"
        /db_xref="taxon:5759"
        /clone_11b="Entamoeba histolytica Sheared DNA"
        /note="vector: pHOSt; Site: 1; Bst I; Constructed at The
        Institute for Genomic Research (TIGR), Rockville, MD.
        Genomic DNA isolated from broth cultures of E. histolytica
        using a method described by Clark and Diamond (Clark,
        C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
        method for isolate identification. Exp. Parasitol.
        77:450.). The DNA was mechanically sheared to give a
        tight size distribution (~2 kb). The v + i method used for
        the library construction is described in detail in Smith,
        H.O. and Venter, J.C. (Making small insert libraries for
        whole genome shotgun sequencing projects. In Genome
        Sequencing: A Practical Approach, eds. M. Vaubin and B.
        Barrell, Oxford University Press, 1999).
BASE COUNT
  185 a 224 c 92 g 415 t
ORIGIN
  Query Match 5.8%; Score 44.4; DB 17; Length 916;
  Best Local Similarity 46.5%; Pred. No. 1.5;
  Matches 144; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 105 CTATGATAAATCACCAGAAATCAACAAGCTATTGATGTCATTTGCTGCTATTGCA 164
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 CGATGATGAAGAGATGAAGAGATGATGAAGAGATGATGAAGAGATGAAGAGATGA 615
QY 165 ACAATCCGAACATAGATCCAAATGAAAGTACTGATGCTGCGATTAATTCGAACGTCA 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 614 CGATGATGAAGAGATGAAGATGAAGATGATGATGCTTGAAGATGATGATTTGACAT 555
QY 225 TGTGATATGATGATTTCAAGGTGAATTAAGCATGCGAAACATTTAGAGCTCGAGATT 284
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 554 CGACATGATGATGATGATTTGATGATGAAGATGAAGATGAAGATGATGAAGATGA 495
QY 285 GAAACAATGAAGAGTCAAGGTGATGCTAATGTCAAAGGTGAAGAGGATTTGTTAAAGC 344
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 CGATGAAGATGAAGATGATGATGATGAAGATGATGAAGATGATGAAGATGAAGATGA 435

```

```

QY 345 TCATTGTTGATGGTTCACATGATATCTGCTGATGAAATGATGATTTAGCATACAA 404
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 TGATGATGAGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375
QY 405 ATGGGAT 414
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 AGACGATGAT 365

RESULT 16
LOCUS CDS00EJ4
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL069257
VERSION AL069257.1 GI:4949400
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazuhiro Osoegawa and
            Aaron Hammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            pl and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
            source
            1..1101
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="BACR29K22"
            /clone_1lb="RPCI-98"
            /note="end : 77"

BASE COUNT 434 a 140 c 25 g 377 t 125 others
ORIGIN
Query Match 5.8%; Score 44.4; DB 17; Length 1101;
Best Local Similarity 38.3%; Pred. No. 1.5;
Matches 125; Conservative 35; Mismatches 166; Indels 0; Gaps 0;

QY 396 AGCATACAATGGGATCTTCATCCACCATCATGATTTGGATATTCAGATT 455
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 AMCHCATTAATAATWTTTTCACAAAATAATCTTTTTCCTTTTTHTTTTTT 822
QY 456 TGTGTGCGCTTGCCCTGGAATTTCTGATGAAGTAACATAGCATCTTTGA 515
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 HTHTTAATAATWAAATTTTAAATAAAAAAATTTWAAAAAATTTTAA 882
QY 516 AGTAGCAATATGCTATGTCACCATATGGTGTCTTCAATCTTGAATCCAT 575
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 AAATTTATTTTAAATWTAATWTAATAATTTTWTWTAATWTAATTTAAW 942
QY 576 TTTTGGCGTTTATGATGATGATGACCGTATTTTCCAGACCGCTAGTAAGAAAT 635
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 AAAATTTGMAAAWATTTTAAATTTTAAATTTTAAWGTAGAAATWAAAT 1002

```

```

QY 636 GACCAAGATATGACACACCATTTAAAGCTGAATTTGAAAAAATTAACCAATAGACAT 695
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 TWTTKTTTTTMAAAAAAARATKTAARKTGAGTTKGAAAAAGRTTMAAGTTAAAT 1062
QY 696 CATTTTCCACATGCTACATCTCTAT 721
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 WGTTTTTTAAAGRAATWTTTATTAT 1088

RESULT 17
LOCUS A0076852
DEFINITION C17-HSP-2361M9.TR C17-HSP Homo sapiens genomic clone 2361M9, DNA
            sequence.
ACCESSION A0076852
VERSION A0076852.1 GI:3438036
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 292)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mhadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES
            source
            1..292
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="2361M9"
            /clone_1lb="C17-HSP"
            /sex="Male"
            /cell_type="Sperm"
            /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
            HindIII"

BASE COUNT 91 a 54 c 43 g 104 t
ORIGIN
Query Match 5.8%; Score 44.2; DB 17; Length 292;
Best Local Similarity 49.8%; Pred. No. 1.6;
Matches 112; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 414 TCTTCATCCACACCATCATGATTTGGATATTCAGATTTTGTGTCCTTGCCCT 473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 TCTTCACCCACGACATGGAACACCAATCTTTTGTGATACATAATTTTATTTTGA 110
QY 474 TGAATTTCTGATGAGAGTAACATTAACATGACATCTTTGAGATGACGACATTCGCTAA 533
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 AGAATTAATCTTAAGTAAATCCTAGAAATATGATATCTTTGTTCTGACACTTTTCATTTA 170
QY 534 TGTTCACACCATATGTTGTTCTTTCATCTTGATCCAAATTTTGGCGTTTATCTGA 593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 GCATAGCAAAACATATGAGAGTCTATCCATACCTTGACCTGTAACAGAGATAGTATTT 230
QY 594 TGTATTACCGCTATTTTCCAGACACCGTACGTAAGGAATGAC 638
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 TTGTTAATTGACAAATTAATGATGCAATTAATGAAAGTAATAATTAC 275

```

REFERENCE	TITLE	COMMENT
LOCUS	BM165203	
DEFINITION	BM165203	
ACCESSION	BM165203	
VERSION	BM165203.1	
KEYWORDS	EST.	
SOURCE	Plasmodium yoelii yoelii.	
ORGANISM	Plasmodium yoelii yoelii.	
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
JOURNAL	1 (bases 1 to 341)	
COMMENT	Carlton, J. M., Daly, T. M., Long, C. A., Bergman, L. W., Valdivia, A. B., Fraser, C. M., and Carucci, D. J.	
	Plasmodium yoelii EST project at TIGR	
	Unpublished (2001)	
	Contact: Jane Carlton	
	Parasite Genomics Group	
	The Institute for Genomic Research	
	9712 Medical Center Drive, Rockville, MD 20850, USA	
	Tel: 301-530-9319	
	Fax: 301-838-0208	
	Email: carltonetigr.org	
	For clone info, please contact the Malaria Research and Reference	
	Reagent Resource Center, ATCC	
	http://www.malaria.mr4.org/mr4pages/index.html	
	Seq primer: ADF.	

BASE COUNT	170 a	39 c	65 g	67 t
ORIGIN				
Query Match		5.8%	Score 43.8;	DB 13; Length 341;
Best Local Similarity	47.9%		Pred. No. 2;	
Matches 126; Conservative		0; Mismatches 137;	Indels 0;	Gaps 0;
OY	110	ATAAAATCACCAGAGAAATCAACAAGCTATTGATGATGCCATTGCTGCTATTGAAACAAT	169	
Db	11	ATAAAGAAACCGATAGCATATATAAGGTGGTGATATAGAACTGTATATAGAGTAGAGA	70	
OY	170	CGGAACATATGATGCCATGAAGTACTGATCATGCCGATTAATTCGAACGTACTGTTG	229	
Db	71	ACGGAGAAAAAGAAATTAATGACTGATCTAATAAAGATGCAGAAAATATTACCAACTGAAG	130	
OY	230	GTATTGTGATTTCAAGAGCTGAATTAGCCATGCCAACAATTGAGGCTCGAGAGATTGAAC	289	
Db	131	AACCTTGATGGTACTTAAATAATTAATTAATAATTCGAATTTGAAAGAAAAAGAAATTAAC	190	

QY	290	AAATGAACGCTACAGGTATGCTATATGCAAGGTGAAGSAGGCTATTGTGAAGCTCAAT	349
Db	191	AAACACACTTCTCCCTCAAAAAGATGTATCACAAAAATTAATATTTGGTATGATTAACA	250
QY	350	TGTTGATCGGTGTTCACGATCAT	372
Db	251	TAACTATGATATCTTAACCAAGAT	273
RESULT	19		
LOCUS	BM160594	456 bp	mRNA linear EST 04-DEC-2001
DEFINITION	EST563117 PyBS Plasmodium yoelii yoelii cDNA clone pYCU30 5' end,		
ACCESSION	BM160594		
VERSION	BM160594.1		
KEYWORDS	EST.		
SOURCE	Plasmodium yoelii yoelii.		
ORGANISM	Plasmodium yoelii yoelii.		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 456)		
TITLE	Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdivya,A.B.,		
JOURNAL	Fraser,C.M. and Carucci,D.J.		
COMMENT	Plasmodium yoelii EST project at TIGR		
	Unpublished (2001)		
	Contact: Jane Carlton		
	Parasite Genomics Group		
	The Institute for Genomic Research		
	9712 Medical Center Drive, Rockville, MD 20850, USA		
	Tel: 301-530-9319		
	Fax: 301-838-0208		
	Email: carlton@tigr.org		
	For clone info, please contact the Malaria Research and Reference		
	Reagent Resource Center, ATCC		
	http://www.malaria.mr4.org/mr4pages/index.html		
	Seq primer: ADF.		

BASE COUNT	249 a	31 c	84 g	92 t
ORIGIN				
Query Match		5.8%	Score 43.8;	DB 13; Length 456;
Best Local Similarity		47.9%	Pred. No. 2;	
Matches 126; Conservative		0;	Mismatches 137;	Indels 0; Gaps 0;
Oy	110	ATAAATACCGAGAAATCAACAAGCTATTGATGATGCCATTGCTCTATTGACAAT	169	
Db	192	ATAAAGAAACCGATGTGCATATATAAAGCTGGAGATATGAACCTTTATGTAGCTAGAGA	251	

RESULT 27
LOCUS BJ371247/c 563 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ371247 Dictyostelium discoidium cDNA library, CF Dictyostelium discoidium cDNA clone ddc57a08 5', mRNA sequence.
ACCESSION BJ371247 GI:19280630
VERSION BJ371247.1
KEYWORDS EST.
SOURCE Dictyostelium discoidium.
ORGANISM Dictyostelium discoidium.
REFERENCE 1 (bases 1 to 563)
AUTHORS Urushinara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoidium at the culmination stage

JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshiniegenes.nig.ac.jp.

FEATURES
source Location/Qualifiers
1..563
/organism="Dictyostelium discoidium"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc57a08"
/clone_1lb="Dictyostelium discoidium cDNA library, CF"
/sex="mat A"
/dev_stage="Culmination stage"

BASE COUNT 276 a 156 c 42 g 88 t 1 others
ORIGIN

Query Match 5.6%; Score 42.8; DB 13; Length 563;
Best Local Similarity 43.6%; Pred. No. 3.6;
Matches 188; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

QY 292 ATGAACGCAAGCGTAGCTGATGCTAATGCAAGGAGAGCGTATTTAAAGCTCATTTG 351
DB 548 AACTAATAGCTTCTGCTGTTGTTGTTGATGTTGTTGCTGCTGTTGTTGATGTT 489
QY 352 TTGATCGTTCACGATGATATCGTCTGATGGAATGATTTAGCATACAAATGGGT 411
DB 488 GTTGTGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 429
QY 412 GATCTTCATCCAAACACTCATGTCATTTGCGAATATCAAGATTTGTTGCTGCTGTC 471
DB 428 GTTGCTGAATGAGAGCTGAGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 369
QY 472 CTGGAATTTCTGATGAGGTACATACATACATACATCTTTGAAGTACGACAAATGCGT 531
DB 368 GTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 309
QY 532 AATGTTGTCACCAATATGTTGCTGTTTCAATCTTGATCCAAATTTTGGCGTTTATCT 591
DB 308 TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 249
QY 592 GATGTAATGACGCTATTTTCCAAACACCGTAGCGTAGAATAATGACCAAAATTTGGCA 651
DB 248 GTTGTAATGTTGAATTTGTTGCTGATTTGTTGTTGTTGTTGTTGTTGTTGTTG 189
QY 652 CCAGCAATTTAAGGTGAATGAAAAAATTACCAATAGACATCTTTTCCAACTGTA 711
DB 188 AACGATTTGCTGAGTTGATGTTGATTTAGATTACACCTTGTGACTGTCGA 129
QY 712 CAATCTTAT 722
DB 128 CAATATTCATT 118

RESULT 28
LOCUS BJ369346/c 483 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ369346 Dictyostelium discoidium cDNA library, CF Dictyostelium discoidium cDNA clone ddc50p01 5', mRNA sequence.
ACCESSION BJ369346 GI:19278729
VERSION BJ369346.1
KEYWORDS EST.
SOURCE Dictyostelium discoidium.
ORGANISM Dictyostelium discoidium.
REFERENCE 1 (bases 1 to 483)
AUTHORS Urushinara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoidium at the culmination stage

JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshiniegenes.nig.ac.jp.

FEATURES
source Location/Qualifiers
1..483
/organism="Dictyostelium discoidium"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc50p01"
/clone_1lb="Dictyostelium discoidium cDNA library, CF"
/sex="mat A"
/dev_stage="Culmination stage"

BASE COUNT 220 a 142 c 25 g 96 t
ORIGIN

Query Match 5.6%; Score 42.6; DB 13; Length 483;
Best Local Similarity 46.6%; Pred. No. 4.1;
Matches 170; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

QY 178 ATGATCCATGAAGTAGTACTGATCATGCGCATTAATTCGAACGTCATGTTGTTG 237
DB 382 ATATTACGTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 323
QY 238 GATTTCAAGGTGAATTAACCATGCGAACATTGAGCTCGAGATTGAACAATGAAA 297
DB 322 CTAATGGAAGTTGATTAACATATATGATGATGTTGAGAGTGTATATGATGAAA 263
QY 298 CGTCAAGGTGATCTAATGTCAAAGGTGAAGAGGATTTTAAGCTCATTTGTTGATC 357
DB 262 TATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 203
QY 358 GGTGTCACGATGATATGCTGCTGATGATATGATTAGCATACAAATTTGGGTGATCT 417
DB 202 GTATATTACACATATATTGCTGCTGACGCTAAAAAATGTTGGAGATTTGTTG 143
QY 418 CATCCAAACATCATCTCATTTTCGATTTCAAGATTTGTTGTTGCCCTTGCCCTGAA 477
DB 142 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 84
QY 478 ATTTGATGAGAGTTAATACATAGCATCTTTTGAAGTACGACAAATTTGCTAATGTT 537
DB 83 ATATATGATGAGTGGAGATTTGATATTACAAAGTTTGCAGAGTACGACATGTTGTTG 24
QY 538 GTCAA 542
DB 23 GAAAA 19

RESULT 29
LOCUS BJ370579/c 484 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ370579 Dictyostelium discoidium cDNA library, CF Dictyostelium discoidium cDNA clone ddc54k15 5', mRNA sequence.

[illegible][illegible]

[illegible]

```

FEATURES                                     Fax: 81-559-81-6855
source                                       Email: tshin@genes.nig.ac.jp.
                                           Location/Qualifiers
                                           1..516
      /organism="Dictyostelium discoideum"
      /strain="AX4"
      /db_xref="taxon:44689"
      /clone_id="dda0013"
      /clone_lib="Dictyostelium discoideum cDNA library, Af"
      /sex="mat A"
      /dev_stage="Aggregation stage"

BASE COUNT          242 a         154 c         27 g         93 t

ORIGIN

Query Match              5.6% Score 42.6; DB 13; Length 516;
Best Local Similarity   46.6%; Pred.No. 4.1;
Matches 170; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

QY    178 ATAGATCCAAATGAAAAGTACCTGCATCATGCCATAAATTCGAACGTCATGTGGATTGTG 237
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     369 ATATTACGTTTGCGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 310
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    238 GATTTCGAAGGTGAATTTAGCCATGCGAAACAATTTAGCCTGCAGATTTCAAAACAATGAAA 297
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     309 CTAATGGAAGTTGATTAACATATATATGTATCTGATTTGGAGAAGTGTTGTAATTTGATGAAA 250
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    298 CGTCAAGGTGATGCTAATGTCAAAGGTGAAGAGGTATTTGTAAGCTCATTTGTGTGATC 357
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     249 TATTGTGGTGGATGATGCTTGGAAATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 190
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    358 GGTGTTCACGATGATATGTCTCGATGGAATATGATTTTAGCATACAAATTTGGTGATCTT 417
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     189 GTATATTACCACATATTATGTGCTGACGTAAAAAAGATGTGGAGATNTGGGTTTTGTG 130
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    418 CATCCAACACCMCATGTCATTTGCGATTTTCAGATTTTTCAGATTTTGGTTCCTTGCCCTTGAA 477
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     129 TTGTGTTGTTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 71
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    478 ATTTCGTGTAAGGTAACATACATGACATCTTTTGAAGTACGACAATTCGCTAATGTT 537
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     70 AATATATGATGGTGGAATTTGATCATTTTCAAAAGTTTGCAGAGTAGACATGGTGTGTGT 11
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    538 GTCAA 542
      |||
      Db     10 GAAA 6

RESULT 33
BU368463/c BU368463 519 bp mRNA linear EST 08-MAR-2002
LOCUS       BU368463 Dictyostelium discoideum cDNA library, CF Dictyostelium
DEFINITION dictioideum cDNA clone ddc6bh19 5', mRNA sequence.
ACCESSION   BU368463
KEYWORDS    BU368463.1 GI:19277846
SOURCE      EST.
            Dictyostelium discoideum.
ORGANISM    Dictyostelium discoideum
REFERENCE   1 (bases 1 to 519)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shn-I,T.
TITLE       Full length cDNA of Dictyostelium discoideum at the culmination
           stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
           Center For Genetic Resource Information
           National Institute Of Genetics
           111 Yata, Mishima, Shizuoka 411-8540, Japan
           Tel: 81-559-81-6856
           Fax: 81-559-81-6855
           Email: tshin@genes.nig.ac.jp.
FEATURES
source
/organism="Dictyostelium discoideum"

```

```

/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc46h19"
/clone_1lb="Dictyostelium discoideum cDNA library, CF"
/dev_stage="Culmination stage"
/sex="mat A"
BASE COUNT      238 a      155 c      27 g      99 t
ORIGIN

Query Match      5.6%; Score 42.6; DB 13; Length 519;
Best Local Similarity 46.6%; Pred. No. 4.1;
Matches 170; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

QY 178 ATGATCCATGAAGAACTGATGATCATGCGATTAATTCAGAGCTCATGTTGGTATG 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 ATATTACGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 319

QY 238 GATTTCAGAGGTGAATTAGCCATGCGAAGCATTCGAGCTCGAGGATTCGAACAATGAAA 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 CTAAATGGAAGTGAATTAACTATATATGATGCTATTGGAGTGGTGAATATGATGAAA 259

QY 298 CGTCAGGTATGCTAATGTCAGAGGTGAAGAGGCTATTGTTAAAGCTCATTTGTTGATC 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 TATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 139

QY 358 GGTGTCAGCATGATATGCTGCTGATGATGATGATTTAGCATACAAATTCGGGATCTT 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 GTATTATACCATATATGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 139

QY 418 CATCAACCATCATGATGATTTGGATTCAGAGATTTGTTGGCTTGTCCCTTGAA 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 TTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 80

QY 478 ATTTCGATGAAGTGAATCAATGACATCTTTTGAATAGACATTCGCTAATGTT 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 ATATATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 20

QY 538 GTCAA 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19 GAAA 15

RESULT 34
BI594832      525 bp      mRNA      linear      EST 07-SEP-2001
LOCUS      BI594832/c
DEFINITION      As_t92_44G04_SKPL Ascaris suum adult male testis germinal zone from
                Alan Scott Ascaris suum cDNA clone As_t92_44G04 5', mRNA sequence.
ACCESSION      BI594832
VERSION      BI594832.1 GI:15498319
KEYWORDS      EST.
SOURCE      pig roundworm.
ORGANISM      Ascaris suum.
                Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
                ; Ascarididae; Ascaris.
REFERENCE      1 (bases 1 to 525)
AUTHORS      Blaxter,M.L., Parkinson,J., Whittton,C., Daub,J., Guiliano,D., Hall
                ,N., Quayle,M. and Barrell,B.
TITLE      Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL      Unpublished (2000)
COMMENT      Contact: Blaxter ML
                Institute of Cell, Animal and Population Biology
                University of Edinburgh
                Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
                3JT, UK.
                Tel: +44 131 650 6760
                Fax: +44 131 670 5450
                Email: mark.blaxter@ed.ac.uk
                The library was prepared by Michelle Lizotte-Waniewski for Alan
                Scott, Johns Hopkins University Medical School, Baltimore..
                Sequencing was performed by the Pathogen Sequencing Unit, Sanger
                Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
                sequence contained a Polya tail (trimmed)
                PCR Primers

```

```

FORWARD: T3
BACKWARD: T7PL
Plates: 44 row: G column: 04
Seq primer: SKPL
High quality sequence stop: 502.
Location/Qualifiers
1. 525
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_t92_44G04"
/clone_1lb="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/Note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT      193 a      149 c      66 g      117 t
ORIGIN

Query Match      5.6%; Score 42.6; DB 13; Length 525;
Best Local Similarity 48.2%; Pred. No. 4.1;
Matches 149; Conservative 0; Mismatches 159; Indels 1; Gaps 1;

QY 180 AGATCCATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 AGTTGCGATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 379

QY 240 TTTCAGAGGTGAATTAGCCATGCGAAGCATTCGAGCTCGAGATTCGAACAATGAAAACG 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 TGTGGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 319

QY 300 TCAAGGTGATCTAATGTCAGAGGTGAAGAGGCTATTGTTAAAGCTCATTTGTTGATCG 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 TGCTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 259

QY 360 TGTTCAGATGATATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 TAAAGCTACTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 139

QY 420 TCCACACTCATGATGATTCGATATTCGAATTCGAATTCGATTCGATTCGATTCGATTCG 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 TCAAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 139

QY 479 TTTCTGATG 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 TTACTGTTG 130

RESULT 35
BJ367414/c      533 bp      mRNA      linear      EST 08-MAR-2002
LOCUS      BJ367414
DEFINITION      dictyostelium discoideum cDNA library, CF Dictyostelium
                discoideum cDNA clone ddc42K13 5', mRNA sequence.
ACCESSION      BJ367414
VERSION      BJ367414.1 GI:19276716
KEYWORDS      EST.
SOURCE      Dictyostelium discoideum.
ORGANISM      Dictyostelium discoideum.
                Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE      1 (bases 1 to 533)
AUTHORS      Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE      Full length cDNA of Dictyostelium discoideum at the culmination
                stage
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasi Shin-I
                Center For Genetic Resource Information
                National Institute of Genetics
                1111 Yata, Mishima, Shizuoka 411-8540, Japan
                Tel: 81-559-81-6856
                Fax: 81-559-81-6855

```


[illegible][illegible]

[illegible]

RESULT 42					
BJ328350/c					
LOCUS	BJ328350	567 bp	mRNA	linear	EST 05-MAR-2002
DEFINITION	BJ328350 Dictyostelium discoidium cDNA library, AF Dictyostelium				
ACCESSION	discoidium cDNA clone dda23124 5', mRNA sequence.				
VERSION	BJ328350				
KEYWORDS	BJ328350.1	GI:19158480			
SOURCE	EST.				
ORGANISM	Dictyostelium discoidium.				
REFERENCE	Dictyostelium discoidium				
AUTHORS	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.				
TITLE	(bases 1 to 567)				
	Unshihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.				
	Full length cDNA of Dictyostelium discoidium at the aggregation				

**JOURNAL
COMMENT**
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES	
source	Location/Qualifiers
	1. 567
	/organism="Dictyostellium discoideum"
	/strain="AX4"
	/db_xref="taxon:44689"
	/clone="gd423124"
	/clone_1ib="Dictyostellium discoideum cdna library, AF"
	/sex="mat A"
	/dev_stage="Aggregation stage"
BASE COUNT	227 a 154 c 47 g 139 t
ORIGIN	

Query Match	5.68;	Score 42.6;	DB 13;	Length 567;
Best Local Similarity	46.68;	Pred. No. 4.1;		
Matches 170; Conservative	0;	Mismatches 194;	Indels 1;	Gaps 1;

QY	178	ATATCATCAATGAAGAGTACCTGATCATGCCATTAATTCGAAAGCTGTTGGATTGTG	237
Db	526	ATATACGTTTGGGCGTTGTTGTTGTGTATGATTTTGTGTAGCTGTGTGTGTGTG	467
QY	238	GATTCCAAAGGTGAATTAGCCATCGAACAATTGAGCTCGAGGATTGAACAACATGAAA	297
Db	466	CTATTCGAAGTTGATTACATATATGATTCCTATTGGAGTGGTTGTAATATTGATBAAA	407
QY	298	CGTCAGGTGATGCTTAATGTCCAAAGGTGAAGAGGGTATTGTTAAAGCTCATTTGGTATC	357
Db	406	TATTTGGTTGATGTTGTGGAATATGTTGCTATGGTGTGATTTGATGCTGTTGTTG	347
QY	358	GGTGTTCACGATGATATTCGTCGATGGAAATATGATTAGCATCAAAATTGGGTGATCTT	417

[illegible]

RESULT 43	
Bj370337/c	
LOCUS	574 bp mRNA linear EST 08-MAR-2002
DEFINITION	Bj370337 Dictyostelium discoideum cDNA library, CF Dictyostelium
	discoideum cDNA clone ddc53d20 5', mRNA sequence.
ACCESSION	Bj370337
VERSION	Bj370337.1 GI:19279720
KEYWORDS	EST.
SOURCE	Dictyostelium discoideum.
ORGANISM	Dictyostelium discoideum
	Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE	1 (bases 1 to 574)
AUTHORS	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE	Full length cDNA of Dictyostelium discoideum at the culmination stage

**JOURNAL
COMMENT**

Unpublished (2002)
Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yatae, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES	
source	Location/Qualifiers
	1. 574
	/organism="Dictyostellium discoideum"
	/strain="Ax4"
	/db_xref="taxon:44689"
	/clone="ddc53d20"
	/clone_lib="Dictyostellium discoideum cdna library, CP"
	/sex="mat A"
	/dev_stage="Culmination stage"
BASE COUNT	259 a 163 c 32 g 120 t
ORIGIN	

Query Match	5.68;	Score 42.6;	DB 13;	Length 574;
Best Local Similarity	46.68;	Pred. No. 4.1;		
Matches 170; Conservative	0;	Mismatches 194;	Indels 1;	Gaps 1;

QY	178	ATGATCCAAATGAAGAACTACCGATCATCGCGATTAATTCGAACTCATGTTGGTATGTG	233
Db	401	ATATTACCTTTGGTGGTTGTGTGTAAGTATGTTTGTGACTGTGTTGGTTGGTGC	3422
QY	238	GATTTCAAAGGTGAATTAAGCCATCCGAAACATTCAGCCTGAGAGTTGAACAACAAATCAA	2977
Db	341	CTAATGGAAGTGTATTAACATATATATGATCCTATTGGAGTGGTGTAAATTTGATCAAA	2822
QY	298	CGTCAAGGTGATGCTAATGTTCMAAGGTGAAGAGGGATTGTAACCTCATTTGTGATC	3577
Db	281	TATTTGGTGGTATGGTTGTGGAATATGTTGGATGTTGCTGATTCGATCGTGTGTCG	2222
QY	358	GGGTTCACGATGATATTCGTCGATCGATGAATATGATTTAGCATCAAAATTTGGGGATCTT	4177
Db	221	GTAATATACCACTATATTGCTGCTGACGCTAAAAAAGATGTTGGAGATTTGGGTTTTGTG	1622
QY	418	CATCCAAACCACTCATGTCAATTCGATATTCMAAGATTTTGTGCTTGTGCCCTTGAA	4777
Db	161	TTTCTTTGTTGTTGTAGTTGTTGTTGTTGCTGCTGTTGTTGTTGTTGTTGCTAAATTTG	10303

OY	478	ATTTCGTGAAGGTAACATAAATGCATCCTTTTGAAGTAGCACCAFTGGCTAAATGTT	537
Dd	102	ATATATGATGGTGAGATGTTACATTTTACAAGAATTGCCAGACTGACACATGGTGTCTTGT	43
OY	538	GTCAA 542	
Dd	42	GAAA 38	
RESULT 44			
LOCUS	BJ335809/c	582 bp	mRNA linear EST 05-MAR-2002
DEFINITION	Bj335809 Dictyostelium discoideum cDNA library, Af Dictyostelium		
ACCESSION	discoidem cDNA clone ddas1j17 5', mRNA sequence.		
VERSION	Bj335809		
KEYWORDS	Bj335809.1 GI:19165939		
SOURCE	EST.		
ORGANISM	Dictyostelium discoideum.		
REFERENCE	Dictyostelium discoideum.		
AUTHORS	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.		
TITLE	1 (bases 1 to 582)		
JOURNAL	Urushihara,H., Tanaka,Y., Kohera,Y., and Shin-i,T. Full length cDNA of Dictyostelium discoideum at the aggregation stage Unpublished (2002) Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-359-81-6855 Email: tshin@genes.nig.ac.jp. Location/Qualifiers 1..582		
FEATURES	/organism="Dictyostelium discoideum" /strain="AX4" /db_xref="taxon:44689" /clone="dda51j17" /clone_lib="Dictyostelium discoideum cDNA library, Af" /sex="mat A" /dev_stage="Aggregation stage"		
BASE COUNT	263 a 159 c 27 g 133 t		
ORIGIN			
Query Match	5.6%; Score 42.6; DB 13; Length 582;		
Best Local Similarity	46.6%; Pred: No. 4.1;		
Matches 170; Conservative	0; Mismatches 194; Indels 1; Gaps 14;		
OY	178	ATAGATCCAAATGAAGTACCATCATCATGCGCGCAATTAATTCGACGTCATGTGTAATGGG	237
Dd	411	ATATTACGTTTGCGTGTGTTGTTGTAGTGTGATTTTGTGTGTGTGTGTGTGTGTGTGTG	352
OY	238	GATTTCAAAGGTGAATTAGCCATGGCAACATTTGAGGCTCGAGATTGGAACAAATGANA	297
Dd	351	CTTAATGAAGTGTGATTAACATATATGATATCGATTGGAGGTGTTGATATATTGATGANA	292
OY	298	CGTCAAGGATGATGCTAAATGTCCAAGGTGAAGGCGATTTGTTAAAGCTATTTGTTGATC	357
Dd	291	TATTTGGTTGGTATGCGTTGGAAATATGTGGTATATGTTGTTGATTTGATTTGATGCTGTG	232
OY	356	GGTGTTCACGATGATATCGTCTCGAATGGAATTTGATTAACAATTTGGTGATCTT	417
Dd	231	GTAATATTACACATATATCTCTGCTGACGTAATAAAAAAGATGTGGACATGTGGGTTTTGTG	172
OY	418	CATCCAACCATCATGTCATTTTGGATATTCAGATTTTGTGTGCCCTGTCCCTGAA	477
Dd	171	TTTGTGTTGTTGTGTGTATGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	113
OY	478	ATTTCGTGAAGGTAACATAAATGCATCCTTTTGAAGTAGCACCAATTCGCTAAATGTT	537
Dd	112	ATATATGATGGTGAGATTTTACATTTTACAAGAATTGCCAGACTGACACATGGTGTCTTGT	53

OY	538	GTCAA	542
Db	52	GAAAA	48
RESULT 45			
BJ366050/c			
LOCUS			
DEFINITION	BJ366050 Dictyostelium discoidium	582 bp	mRNA linear EST 08-MAR-2002
ACCESSION	discoidium cDNA clone ddc37g16 5', mRNA sequence.		
VERSION	BJ366050		
KEYWORDS	BJ366050.1 GI:19275352		
SOURCE	EST.		
ORGANISM	Dictyostelium discoidium.		
REFERENCE	Dictyostelium discoidium.		
AUTHORS	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.		
TITLE	1 (bases 1 to 582)		
	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.		
	Full length cDNA of Dictyostelium discoidium at the culmination stage		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: Tadasu Shin-i		
	Center for Genetic Resource Information		
	National Institute of Genetics		
	111 Yata, Mishima, Shizuoka 411-8540, Japan		
	Tel: 81-559-81-6856		
	Fax: 81-559-81-6855		
	Email: tshin@genes.nig.ac.jp.		
FEATURES	Location/Qualifiers		
source	1..582		
	/organism="Dictyostelium discoidium"		
	/strain="AX4"		
	/db_xref="taxon:44689"		
	/clone="ddc37g16"		
	/clone.lib="Dictyostelium discoidium cDNA library, CF"		
	/sex="mat A"		
	/dev_stage="Culmination stage"		
BASE COUNT	271 a 161 c 30 g 120 t		
ORIGIN			
Query Match	5.6%; Score 42.6; DB 13; Length 582;		
Best Local Similarity	46.6%; Pred. No. 4.1;		
Matches 170; Conservative	0; Mismatches 194; Indels 1; Gaps 1;		
OY	178	ATAGATCCAAATGAAGTACCTGATCATGCCGTAATAATCGAAGCGTCAGTTGGATTGTG	237
Db	378	AAATTACGTTTGCGTGTGTTGTTACTGATATTTGTTGAGTGTGTTGTTGTTG	319
OY	238	GATTTCAAGGTGAATTAGCCATGCGGAACATTGAGGCTCGAGATTGAAAACAATGAA	297
Db	318	CYAAAGAAGTGCATTACATATATATGATTCGATTTGGAGAGTGCTGTAATATTGATGAA	259
OY	298	CGTCAGGTCGATGCATTAAGTCCAAGGTGAAGGGTATTTGTTAAAGTCATTTGTTGATC	357
Db	238	TATTTGGTTGGTATGGTTGGAAATATTTGGTATNGTTGATTGATGATGCTGTTGTG	199
OY	358	GGTTCACGATGATATGCTCGATGGAATATGATTAGCATACAAATGGGTCATCTT	417
Db	198	GTAATTACOCATATATGCTGCTGAGCAAAAAAGATGTTGCAATATGTGGTTTGTG	139
OY	418	CATCCAACCACTCATGTCATTTGGATATTCAGATTTTGTGTTGCTGTCCCTGAA	477
Db	138	TTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG-T-A	80
OY	478	ATTTCGATGAAGGTACACATACATGACATCTTTTGAAGTCAGCAATTCGCTAATGTT	537
Db	79	AAATATGATGATGGAGATTTGACATTTTCAAAAGTTTGCAGAGTGAACATGGTGTGTTGT	20
OY	538	GTCAA	542
Db	19	GAAAA	15

[illegible][illegible]

REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
AUTHORS 1 (bases 1 to 590)
TITLE Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T
Full length cDNA of Dictyostelium discoideum at the culmination stage

JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel.: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES location/Qualifiers
source
1..590
/oranism="Dictyostelium discoideum"
/strain="AXA"
/db_xref="taxon:44689"
/cclone="ddc43g10"
/cclone_id="Dictyostelium discoideum cdna library, CF"
/sex="mat A"
/dev_stage="Culmination stage"

BASE COUNT 273 a 161 c 31 g 125 t

ORIGIN

```

Query Match          5.6%; Score 42.6; DB 13; Length 590;
Best Local Similarity 46.6%; Pred.No.4.1;
Matches 170; Conservative 0; Mismatches 194; Indels 1; Gaps 1;

QY   178 ATAGATCCAAATGAAGTACCTGATCATGCCGATAATTGCAGACGTCAATGGTGATTTGTG 237
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    361 AATTACGTTGCGTGGTGGTGGTGTTCAGTGTGAATTTGGTAGTGTTGGTGGTGGTG 322

QY   238 GATTTCAGAAGTGAATTAATGCCATCGAAGAACATTTGAGGCTCGAGATTGCAACAATGAAA 297
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    321 CTAATGAAGAGTTGATTAACATATATATGATATTCATTTGGAGTGGCTGTATATATGATGAAA 262

QY   298 CGTCAAGGTGATGTCAAATGTCAAAGGTGAAGGGTATTTGTTAAAGCTCATTTGTTGATC 357
      || || || || || | | | | | | | | | | | | | | | | | | | | | |
Db    261 TATTTGGTTGGTATCGTTGGTGAATATGTGGTATATGTTGATTTGATGCTGTTGGTG 202

QY   358 GGTGTTCAAGATGATGTGTCATGATGGAATATGATTAACATAAATGGGTGATCTT 417
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    201 GTATTTTACCACATATATGCTGCTAGCTAAAAAAGATGTTGGAATGTTGGCTTTTGIG 142

QY   418 CATCCAACCACTCATGTCAATTCGGATATTCAGAATTTGTTGTCCTGTGCCCTGAA 477
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    141 TTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTAAT 83

QY   478 ATTTCGTGAATGAAGTGAACATATACATGACATCTTTTGAAGTACGACAAATGCCCTAATGT 537
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    82 ATAATVGAATGTGAGATTGTACATTTTACAAAGTTTGCAGAGTGAGACATGGTGTGTGT 23

QY   538 GTCAA 542
      ||
Db    22 GA AAA 18
```

RESULT 49
Bj368828 592 bp mRNA linear EST 08-MAR-2002
LOCUS Bj368828 Dictyostelium discoideum cdna library, CF Dictyostelium
DEFINITION dictoideum cdna clone ddc48b08 5', mRNA sequence.
ACCESSION B368828
VERSION B368828.1 GI:19278211
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination stage

JOURNAL	Unpublished (2002)					
COMMENT	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.					
FEATURES						
Source	1..592 /organism="Dictyostellium discoideum" /strain="AX4" /db_xref="taxon:44689" /clone="ddc48b08" /clone_lib="Dictyostellium discoideum cDNA library, Cf" /sex="mat A" /dev_stage="Culmination stage"					
BASE COUNT	274 a 161 c 31 g 126 t					
ORIGIN						
Query Match	5.6%; Score 42.6; DB 13; Length 592; Best Local Similarity 46.6%; Pred. No. 4.1; Matches 170; Conservative 0; Mismatches 194; Indels 1; Gaps 1;					
OY	178	ATAGTCCAAATGAAGTAGCTACGTCGCCGATAATTGCAACGCATCGTATGTCTG	237			
DB	384	AATATTACTTTGGTGTTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	325			
OY	238	GATTTCGAAGGTGAATTAAGCATCGAACAATTCAGCCTCGAGATGGAACAATGAAA	297			
DB	324	CTAATGGAAGTTGATTACATATATATGTCGATTTGGGAGTGGTGTATATGATGAAA	265			
OY	298	CGTCAGGTGATGCTAATATGCAAGGTGAAGAAGGTATTTAAAGCTATTGTGTATC	357			
DB	264	TATTTGGTTGGTATGTTGTGTGGAATATGTGTGTATGTGTGTGTGTGTGTGTG	205			
OY	358	GGTGTTCAGATGATATGCTCTCATGTGGAATATGATTTAGCATNCAATTTGGTGATCTT	417			
DB	204	GTATATTCACCATAATATGCTGCTGAGGTAAAAAGATGTTGGAGATGTGGGTTTTGTG	145			
OY	418	CATCCAACCACTCATGTCATTTTGGATATTCACAAGATTTTGGTTGCTTCCCTTGAA	477			
DB	144	TTTGTGTTGTTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATTTG-A	86			
OY	478	ATTTCGTGATGAAGGTATACATAACAATGACATCTTTTGAAGTAGACAAATTCGTATGT	537			
DB	85	ATATATGATGTTGAGATTGTACATTTTCAAAGTTTGCACAGATGAGACATGGTGTGTGT	26			
OY	538	GTCAA 542				
DB	25	GAAAA 21				
RESULT 50						
B0325409/c						
LOCUS	B0325409	593 bp	mRNA	linear	EST 05-MAR-2002	
DEFINITION	B0325409 Dictyostellium discoideum cDNA library, AF Dictyostellium					
ACCESSION	B0325409					
VERSION	B0325409.1 GI:19155539					
KEYWORDS	EST.					
SOURCE	Dictyostellium discoideum.					
ORGANISM	Dictyostellium discoideum					
REFERENCE	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.					
AUTHORS	1 (bases 1 to 593) Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.					
TITLE	Full length cDNA of Dictyostellium discoideum at the aggregation stage					
JOURNAL	Unpublished (2002)					
COMMENT	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan					

Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
Source Location/Qualifiers

1. 593
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda14g1"
/clone_lib="Dictyostellium discoideum cdna library, AF"
/sex="mat A"
/dev_stage="Aggregation stage"

BASE COUNT 274 a 161 c 31 g 127 t
ORIGIN

Query Match 5.68; Score 42.6; DB 13; Length 593;
Best Local Similarity 46.68; Pred. No. 4.1;
Matches 170; Conservative 0; Mismatches 194; Indels 1; Gaps 1;
OY 178 ATAGATCCAAATGAAGTACCTGATCATGCCGATTAATTCGAACGTCATGTTGTAATGTG 237
DB 384 ATATTACGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 325
OY 228 GATTTCAAAGTGAATTAAGCCATGCGAAACATTGAGCTCGAGATTGAAACAAATGAAA 297
DB 324 CTAATGGAAGTTGATTACATATATGTAATGCGATTTGGAGTGTGTTGTAATATTGATGAAA 265
OY 298 CGTCAAGGTGATGCTAATGTCAAAGTGAAGGCTATTGTTAAAGCTCATTGTTGATC 357
DB 264 TATTTGGTTGGTANGGTTGGAATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 205
OY 358 GGTGTTCAAGTATGATGCTGATGATGGAATATGATTAGCATACAATTTGGGTATCTT 417
DB 204 GTATATTACCACTATATGCTGCTGAGCTAAAAAGATGTTGAGATGTTGGGTTTGTG 145
OY 418 CATCAACCACTCATGTCATTTCGATATTCGAAGATTGTTGTTGCTTGCCTTGAA 477
DB 144 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 86
OY 478 ATTTCTGATGAAGTACATACATGACATCTTTGAAGTACGACAAATTCGCTAATGTT 537
DB 85 ATATATGATGTGAGATTGTACATTTCAAAGTTGCGAGCTGAGACATGTTGTTGTTG 26
OY 538 GTCAA 542
DB 25 GAAAA 21

Search completed: February 20, 2003, 21:19:55
Job time : 1519 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 18:26:18 ; Search time 2013 Seconds
(without alignments)
3079.433 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 213
Sequence: 1 MMKFLIAAVFAVAVSADPL.....VRKEMTKVIAPAFKRELEKN 213

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2054640 seqs, 14551402878 residues

Word size: 1
Total number of hits satisfying chosen parameters: 4106479

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-Q=/cgn2.1/USFIO_spool/US10024955/runat_14022003_130735_24251/app-query.fasta.1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIN=200 -THR.SCORE=quality -THR.MIN=1 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NOR=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10024955.cgn.1.1.3745.crunat.14022003_130735_24251 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1:	GenEmbl:*
2:	gb_ba:*
3:	gb_hc9:*
4:	gb_in:*
5:	gb_om:*
6:	gb_ov:*
7:	gb_pat:*
8:	gb_ph:*
9:	gb_pl:*
10:	gb_pr:*
11:	gb_ro:*
12:	gb_sts:*
13:	gb_sy:*
14:	gb_un:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_lnu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pat:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_sts:*
28:	em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rnd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	213	100.0	761	3 S80655	S80655 allergen De
2	213	100.0	761	6 AX028702	AX028702 Sequence
3	25	11.7	812	3 DPU37044	DPU37044 Dermatophag
4	25	11.7	812	6 AR098569	AR098569 Sequence
5	25	11.7	812	6 AX028697	AX028697 Sequence
6	9	4.2	317	6 AX368336	AX368336 Sequence
7	9	4.2	334	11 HS268WE5	266899 H.sapiens D
8	9	4.2	1117	4 AF076282	AF076282 Mustela v
9	9	4.2	1660	4 AF107544	AF107544 Ovis arie
10	9	4.2	21667	2 AC017321	AC017321 Drosophill
11	9	4.2	135792	2 AP003544	AP003544 Oryza sat
12	9	4.2	146738	2 AC126526	AC126526 Rattus no
13	9	4.2	149776	9 AC012070	AC012070 Homo sapi
14	9	4.2	155633	8 AP003344	AP003344 Oryza sat
15	9	4.2	158456	2 AP004731	AP004731 Oryza sat
16	9	4.2	158895	3 AC009345	AC009345 Drosophill
17	9	4.2	166261	9 AC073336	AC073336 Homo sapi
18	9	4.2	167596	2 AC013177	AC013177 Drosophill
19	9	4.2	173476	2 AC117360	AC117360 Rattus no
20	9	4.2	183007	3 AC007257	AC007257 Drosophill
21	9	4.2	189649	3 AC007819	AC007819 Drosophill
22	9	4.2	231732	3 AE003764	AE003764 Drosophill
23	9	4.2	260624	3 AE003625	AE003625 Drosophill
24	8	3.8	24	6 AX028708	AX028708 Sequence
25	8	3.8	205	14 FCSH55R1	M14290 Feline sarc
26	8	3.8	1009	5 AY004870	AY004870 Gallus ga
27	8	3.8	1056	3 AY060711	AY060711 Drosophill
28	8	3.8	1128	6 AX123559	AX123559 Sequence
29	8	3.8	1220	8 AE272757	AE272757 Zea mays
30	8	3.8	1239	6 E05296	E05296 DNA sequenc
31	8	3.8	1324	1 CGRECAG	X77384 C.glutamincu
32	8	3.8	1402	1 CGU14965	U14965 Corynebacte
33	8	3.8	1590	14 AF504294	AF504294 Natrobil s
34	8	3.8	1827	5 AF265671	AF265671 Xenopus l
35	8	3.8	1833	6 AX244529	AX244529 Sequence
36	8	3.8	1965	1 STMHRDC	M90412 Streptomyce
37	8	3.8	2054	5 U01047	U01047 Gallus dome
38	8	3.8	2753	14 FCVGP	K01803 Feline leuc
39	8	3.8	2757	5 AF376334	AF376334 Xenopus l
40	8	3.8	3022	6 AR193470	AR193470 Sequence
41	8	3.8	3023	6 BD006097	BD006097 Sequence
42	8	3.8	3023	6 BD006097	BD006097 Feline po
43	8	3.8	3409	9 AK092132	AK092132 Homo sapi
44	8	3.8	3410	1 MTLHTT	D14071 M. glycogen
45	8	3.8	3410	1 MTLHTT	D14071 M. glycogen

RESULT 1

ALIGNMENTS

S80655
LOCUS S80655 761 bp mRNA linear INV 10-MAY-1996
DEFINITION allergen Der f 7 [Dermatophagoides pteronyssinus-house dust mites, mRNA, 761 nt].
ACCESSION S80655
VERSION S80655.1 GI:1311688
KEYWORDS Dermatophagoides pteronyssinus.
SOURCE Dermatophagoides pteronyssinus.
ORGANISM Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae; Dermatophagoides.
REFERENCE 1 (bases 1 to 761)
AUTHORS Shen,H.D., Chua,K.Y., Lin,W.L., Hsieh,K.H. and Thomas,W.R.
TITLE Molecular cloning and immunological characterization of the house dust mite allergen Der f 7
JOURNAL Clin. Exp. Allergy 25 (10), 1000-1006 (1995)
MEDLINE 96120794
PUBMED 8536554
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI gblseq 174942] from the original journal article.
FEATURES
source Location/Qualifiers
1..761
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
1..761
/gene="allergen Der f 7"
43..684
/gene="allergen Der f 7"
/note="This sequence comes from Fig. 1"
/codon_start=1
/product="allergen Der f 7"
/protein_id="AAB35977.1"
/db_xref="GI:1311689"
/translation="MKKFLIAAVAFVANSADPIHYDKITTEINKAIDDAIAIEQSE
TIDPMKVPDHADKFERHVGIVDFKGLAARNIEARGLKMKRQGDANVGEIGIVAAH
LLIGVHDDIVSMEDLAYKLGDLHPHTHYISIDOPVVALSIEISDEGNTMTSEFVR
QFANVNHIGSLSIDPIFGVLSIDVLTALFQDVRKEMTKVLAFAKRELEKN"
BASE COUNT 245 a 134 c 137 g 245 t
ORIGIN
Alignment Scores:
Pred. No.: 8.8e-226 Length: 761
Score: 213.00 Matches: 213
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 3
US-10-024-955-7 (1-213) x S80655 (1-761)
QY 1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
Db 43 ATGATGAATTTTGTGGATGCTGCGCGTCATTTGTCGCCGTTTCGGCTGATCCAA 102
QY 21 HistTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAlaIle 40
Db 103 CACTATGATTAATTCACCGAAGAATCAACAAGCTATTATATGTCATTCGCTCAT 162
QY 41 GluGlnSerGluThrIleAspPrometLysValProAspHisAlaAspLysPheGluArg 60
Db 163 GAACAATCCCAAAACAATAGTCCAAATGAAGTACCTGATCGATCGCATTAATTCGA 222
QY 61 HisValGlyIleValAspPheLysGlyLysAlaMetArgAsnIleGluAlaArgGly 80
Db 223 CAGGTGGATTTGTCGATTTCAAGGTGAATGAGCCATGCAACATTTGAGGCTCGA 282
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys 100
Db 283 TTGAACAATGAAGCTCAAGTGATGCTAATGTCAAAAGTGAAGGGATTTGTTAAA 342
QY 101 AlAHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120

|||||
Db 343 GCATATTGTTGATCGGTGTTACAGATGATGCTCGATGGAATATGATTACATAC 402
QY 121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValAla 140
Db 403 AAATGGGATGATCTTCATCCACACCATGATGATTTGGATTTCAAGATTTTGGTT 462
QY 141 AlLeuSerLeuGluLeuSerAspGluGlyAsnIleThrMetThrSerPheGluValArg 160
Db 463 GCCTTGCCCTTGAATTTTGATGAAGTAACATGACATCTTTTGAAGTACGA 522
QY 523 CAATCGCTAATGTTGCAACCATATTGCGTCTTCAATCTTGATCCAAATTTTGGC 582
Db 161 GlnPheAlaAsnValAlaAsnHisIleGlyLysSerIleLeuAspProIlePheGly 180
QY 523 CAATCGCTAATGTTGCAACCATATTGCGTCTTCAATCTTGATCCAAATTTTGGC 582
Db 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys 200
QY 583 GTTTATCTGATGATTTGACCGCTATTTTCCAAAGACACCGTACGTAAGAAATGACCAA 642
Db 201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
Db 643 GTATTGCGACACCATTTTAACGTGATTTGGAATAAAT 681
RESULT 2
AX028702 761 bp DNA linear PAT 16-SEP-2000
LOCUS AX028702 Sequence 6 from Patent EP1018550.
DEFINITION AX028702
ACCESSION AX028702
VERSION AX028702.1 GI:10189815
KEYWORDS European house dust mite.
SOURCE Dermatophagoides pteronyssinus
ORGANISM Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae; Dermatophagoides.
REFERENCE 1 (bases 1 to 761)
AUTHORS Thomas,W.R. and Chua,K.Y.
TITLE Allergenic protein and peptides from house dust mite and uses thereof
JOURNAL Patent: EP 1018550-A 6 12-JUL-2000;
INST CHILD HEALTH RESEARCH (AU)
FEATURES
source Location/Qualifiers
1..761
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
43..684
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09235.1"
/db_xref="GI:10189816"
/translation="MKKFLIAAVAFVANSADPIHYDKITTEINKAIDDAIAIEQSE
TIDPMKVPDHADKFERHVGIVDFKGLAARNIEARGLKMKRQGDANVGEIGIVAAH
LLIGVHDDIVSMEDLAYKLGDLHPHTHYISIDOPVVALSIEISDEGNTMTSEFVR
QFANVNHIGSLSIDPIFGVLSIDVLTALFQDVRKEMTKVLAFAKRELEKN"
BASE COUNT 245 a 134 c 137 g 245 t
ORIGIN
Alignment Scores:
Pred. No.: 8.8e-226 Length: 761
Score: 213.00 Matches: 213
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-024-955-7 (1-213) x AX028702 (1-761)
QY 1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
Db 43 ATGATGAATTTTGTGGATGCTGCGCGTCATTTGTCGCCGTTTCGGCTGATCCAA 102
QY 21 HistTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAlaIle 40

Db 103 CACTGATGATAAAATCACCGAAGAAATCAACAAAGTAATTGATGATGCCATTGCTGCTATT 162
QY 41 GUGInsercIurthrIleaspprometIysValproasphIsIaasplysPhegluArg 60
Db 163 GAACAATCCGAACAATAGATCCAAATGAAGTACCTGATCATGCCGTAATATTCGAACGT 222
QY 61 HisValgIyIleValasPheIysgIyGluLeuAlaMetarGasnIleGluAlaArgIy 80
Db 223 CAGTGTGGATTGTGGATTCCAAGGATGATGCCATGCCGAACAATGAGCGCTCGAGGA 282
QY 81 LeuIysGImetIysarGInglIyAspAlaAsnValIysgIyGluGluIyIleValIys 100
Db 283 TTGAAACAATAGAAACGTCAAGGTGATGCTAAATGTCAAAGGTGAAGGCGTATTGTTAAA 342
QY 101 AlaHisLeuLeuIleGlyValHisAspAspIleValSermetIuIyrrAspIleuAlaTyr 120
Db 343 GCTCATTTGTTGATCGGTGTCACGATGATATGCTCGATGGAATATGATTTAGCATAC 402
QY 121 LysLeuGlyAspleuHisprothrThrHisValIleSeraspIleGlnAspPheVal 140
Db 403 AATTTGGGATCTTCATCCAAACCACTCAATGTCATTTGGATATTCAGATTTTGTTGTT 462
QY 141 AlaLeuSerLeuGluIleSeraspGluGlyAsnIleThrMetThrSerPheGluValArg 160
Db 463 GCCTTGCTCCCTGAATTTCTGATGAAGTAACATGAACATGACATCTTTTGAAGTACGA 522
QY 161 GlnPheAlaAsnValIaAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
Db 523 CAATTCGCTPATGTCTCAACCAATATGGTGCTTCATTCATTCGATTCGCAATTTTGGC 582
QY 181 ValLeuSeraspValLeuThrAlaIlePheGlnAspThrValArgIysGImetThrIys 200
Db 583 GTTTATTCGATGTATTGACCGCTATTTCACAGACACCGTACGTAAAGAAATGACCAG 642
QY 201 ValLeuAlaProAlaPheIysarGluLeuGluIysAsn 213
Db 643 GTATTGGCACACGATTAAACGTGAATTGCAAAAAAAT 681
RESULT 3
DPUS37044 812 bp mRNA linear INV 31-OCT-1995
LOCUS Dermatophagoides pteronyssinus Der p 7 allergen polypeptide mRNA,
DEFINITION complete cds.
ACCESSION U37044
VERSION U37044.1 GI:1045601
KEYWORDS
SOURCE Dermatophagoides pteronyssinus.
ORGANISM Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;
Dermatophagoides.
REFERENCE 1 (bases 1 to 812)
AUTHORS Shen,H.D., Chua,K.Y., Lin,K.L., Hsieh,K.H. and Thomas,W.R.
TITLE Molecular cloning of a house dust mite allergen with common
antibody binding specificities with multiple components in mite
extracts
JOURNAL Clin. Exp. Allergy 23 (11), 934-940 (1993)
MEDLINE 20239583
PUBMED 10779281
REFERENCE 2 (bases 1 to 812)
AUTHORS Thomas,W.R.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-1995) Wayne R. Thomas, Molecular Biology
Division, Institute for Child, Health Research, Princess Margaret
Hospital, Roberts Road, Subiaco, Western Australia, 6008, Australia
FEATURES
source 1..812
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
/clone="HD6"
68..715
/codon_start=1
/product="Der p 7 allergen polypeptide"

/protein_id="AAA80264.1"
/db_xref="GI:1045602"
/translation="MKILLIAAAFAVASADPIHDKITFEELINKAVDEAVAIEKSE
TRDPKRVDPDSKFERHIGIIDLKEGLDRNTOVGILKMRVGAANYSESGYVKAH
LIVGVHDDVSNMEYIDATKGLDLPHTHTYISIDQFVLSLEVSEGMWLTSTFVR
QFANVNHIGLSILDPIFAVLSDVLTALFDPTVREMTKVIAPAFKELERNQ"
sig_peptide
mat_peptide
misc_signal
polyA_signal
BASE COUNT 270 a 132 c 143 g 267 t
ORIGIN
Alignment Scores:
Pred. No.: 6,59e-18 Length: 812
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.74% Indels: 0
DB: 3 Gaps: 0
US-10-024-955-7 (1-213) x DPUS37044 (1-812)
QY 155 ThrSerPheGluValArgGlnPheAlaAsnValIaAsnHisIleGlyGlyLeuSerIle 174
Db 530 ACATCGTTGCAAGTACGTCAATTTGCCAATGTTGTCATATATGTTGCTTCAATT 589
QY 175 LeuAspProIlePhe 179
Db 590 TTGGATCCAAATTTTC 604
RESULT 4
AR098569 812 bp DNA linear PAT 14-FEB-2001
LOCUS AR098569
DEFINITION Sequence 1 from patent US 6077517.
ACCESSION AR098569
VERSION AR098569.1 GI:12808335
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 812)
AUTHORS Thomas,W.Robert. and Chua,K.-Y.
TITLE House dust mite allergen, Der p VII, and uses thereof
JOURNAL Patent: US 6077517-A 1 20-JUN-2000;
FEATURES
source 1..812
/organism="unknown"
BASE COUNT 270 a 133 c 142 g 267 t
ORIGIN
Alignment Scores:
Pred. No.: 6,59e-18 Length: 812
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.74% Indels: 0
DB: 6 Gaps: 0
US-10-024-955-7 (1-213) x AR098569 (1-812)
QY 155 ThrSerPheGluValArgGlnPheAlaAsnValIaAsnHisIleGlyGlyLeuSerIle 174
Db 530 ACATCGTTGCAAGTACGTCAATTTGCCAATGTTGTCATATATGTTGCTTCAATT 589
QY 175 LeuAspProIlePhe 179
Db 590 TTGGATCCAAATTTTC 604
RESULT 5

AX028697
LOCUS AX028697 812 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent EP1018550.
ACCESSION AX028697
VERSION AX028697.1 GI:10189810
KEYWORDS
SOURCE
ORGANISM European house dust mite.
Dermatophagoides pteronyssinus
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;
Dermatophagoides.
1 (bases 1 to 812)
Thomas, W.R. and Chua, K.Y.
Allergenic protein and peptides from house dust mite and uses
thereof
Patent: EP 1018550-A 1 12-JUL-2000;
INST CHILH HEALTH RESEARCH (AU)
FEATURES
source
location/Qualifiers
1..812
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
68..715
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09234.1"
/db_xref="GI:10189811"
/translation="MMKLLILIAAFAVNASADPIHYDKITEINKAVDEVAATKESK
TFDPMKVPDHSDFEERHIGIIDLKGLDMNIOVRGLKQKRGADAVKSEGVKAKH
LLVGVHDDVYVSMEDLAYKGLDLPNTHVSDIODPVELSLVSEGGNMTLTSFEVR
QFANVNVNHIGSLIDPPIFAVLSDVLAITQDIYRAEMTVLAPAFKKELRNQG"
mat-peptide 119
BASE COUNT 270 a 133 c 142 g 267 t
ORIGIN

Alignment Scores:
Pred. No.: 6,59e-18 Length: 812
Score: 23.00 Matches: 23
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.74% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x AX028697 (1-812)

QY 155 ThisSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIle 174
|||||
Db 530 ACATCGTTCCGAAGTACGTCATTTGCCAATGTGTGCATCATCATTTGCGCTTTCATT 589
QY 175 LeuAspProlIePhe 179
|||||
Db 590 TTGCATCCCAATTTC 604

RESULT 6
LOCUS AX368336 317 bp DNA linear PAT 16-FEB-2002
DEFINITION Sequence 1046 from Patent WO0204514.
ACCESSION AX368336
VERSION AX368336.1 GI:18856409
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Wang, T., Watanabe, Y., Henderson, R.A., Johnson, J.C., Reter, M.W.,
Marinetakis, M., Carter, D., Fanger, G.R., Vedic, T.S., Bangur, C.S.,
McNabb, A., Fanger, N., Switzer, A., McNeill, P.D. and Clapper, J.D.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0204514-A 1046 17-JAN-2002;
CORIXA CORPORATION (US)
FEATURES
source
location/Qualifiers
1..317

/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 57 a 84 c 110 g 66 t
ORIGIN

Alignment Scores:
Pred. No.: 1.39 Length: 317
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x AX368336 (1-317)

QY 188 AlalIlePheGlnAspThrValArgLys 196
|||||
Db 309 GCCATTTCCAGACGACGCTGAGGAG 283

RESULT 7
LOCUS HS268WE5 334 bp DNA linear STS 04-NOV-1995
DEFINITION H. sapiens DNA segment containing (CA) repeat; clone AFM268we5;
single read, sequence tagged site.
ACCESSION Z66899
VERSION Z66899.1 GI:1051682
KEYWORDS STS; CA repeat; dinucleotide repeat; GT repeat; microsatellite DNA;
microsatellite marker; repeat polymorphism.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS Weissenbach, J.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1995) Genethon, B.P. 60, 91002 Evry Cedex France.
COMMENT E-mail: Jean.Weissenbach@genethon.fr
FEATURES
source
location/Qualifiers
1..334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/cell_line="CEPH 134702"
/clone_lib="genomic DNA"
/note="cloning vector is M13mp18"

BASE COUNT 119 a 71 c 56 g 85 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 1.46 Length: 334
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x HS268WE5 (1-334)

QY 141 AlaleuSerLeuGluIleSerAspGlu 149
|||||
Db 128 GCTCTTCTCTGAGACTCTGATGAA 154

RESULT 8
LOCUS AF076282 1117 bp mRNA linear MAM 09-AUG-1999
DEFINITION Mustela vison clone SP-B/Short similar to surfactant-associated
protein A mRNA sequence.
ACCESSION AF076282
VERSION AF076282.1 GI:5712109
KEYWORDS
SOURCE Mustela vison.
ORGANISM Mustela vison

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae; **gene**
 AUTHORS 1 (bases 1 to 1117) **CDS**
 TITLE Gram-Nielsen, S., Storgaard, T., Broil, S., Christensen, J. and Alexandersen, S.
 JOURNAL Interstitial pneumonia in mink kits: Kinetics of Aleutian mink disease parvovirus (ADV) infection
 REFERENCE Unpublished
 AUTHORS 2 (bases 1 to 1117)
 TITLE Gram-Nielsen, S., Storgaard, T., Broil, S., Christensen, J. and Alexandersen, S.
 JOURNAL Direct Submission
 JOURNAL Submitted (06-JUN-1998) Lab. for Genotechnology, Danish Veterinary Institute for Virus Research, Lindholm, Kalvehave 4771, Denmark
 FEATURES Location/Qualifiers
 source 1..1117
 /organism="Mustela vison"
 /db_xref="taxon:9667"
 /clone="Sp-B/Short"
 /haplotype="A/A"
 /tissue_type="Lung"
 /dev_stage="adult"
 misc_feature 1..1117
 /note="Similar to surfactant-associated protein A"
 BASE COUNT 230 a 330 c 322 g 228 t 7 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.29 Length: 1117
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.238 Indels: 0
 Gaps: 0
 DB: 4
 US-10-024-955-7 (1-213) x AF076282 (1-1117)
 QY 188 AaIlePhcInaSpHrValarglys 196
 LOCUS |||||||||||||||||||||||||
 DB 32 GCCATTTCACGACACAGCGCAAG 58
 RESULT 4
 AF107544 1660 bp mRNA linear MAM 18-JAN-2002
 LOCUS AF107544
 DEFINITION Ovis aries pulmonary surfactant-associated protein B (SP-B) mRNA,
 partial cds.
 ACCESSION AF107544
 VERSION AF107544.1 GI:6492135
 KEYWORDS
 SOURCE Ovis aries.
 ORGANISM Ovis aries
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Ovis.
 TITLE 1 (bases 1 to 1660)
 AUTHORS Pietschmann, S.M. and Pison, U.
 JOURNAL cDNA cloning of ovine pulmonary SP-A, SP-B, and SP-C: isolation of two different sequences for SP-B
 MEDLINE Am. J. Physiol. 278 (4), L765-L778 (2000)
 REFERENCE 20215263
 PUBMED 10749754
 JOURNAL 2 (bases 1 to 1660)
 TITLE Pietschmann, S.M. and Pison, U.
 AUTHORS Direct Submission
 JOURNAL Submitted (18-NOV-1998) Klinik fur Anesthesiologie, Charite, Campus Virchow-Klinikum, Augustenburger Platz 1, Berlin 13353, Germany
 FEATURES Location/Qualifiers
 source 1..1660
 /organism="Ovis aries"
 /db_xref="taxon:9940"
 /clone="Sp-B 34"
 /tissue_type="Lung"

gene 1..1660
 CDS /gene="SP-B"
 <1..1125
 /gene="Sp-B"
 /product="pulmonary surfactant-associated protein B"
 /protein_id="AA14195.1"
 /db_xref="GI:6492136"
 /translation="AKSRLLPWLLLLPMLGCLGSAAGTYSILTCAGPFWCOSILE
 QALQCALGHCLEWGHAEADDDQECENTSRITLTKAKAIFODTVKRLBEDCDV
 LPLKLVPOCRHLLDTPPLIIDHPOSQMNPFTIQHVGCLKPRHPGKPEMPGL
 LDKMALPLPGALQAKPGPOTODSQQEPPIPLPCWLCRTLLIRIQAVIKRGVLA
 VAOYCHVPLLVGLGICOCLEVERYSVILDTLLGRMLPOLVCGGLVLRCSSESPAGPALP
 ALGSLPGEMLPDSDSCOLCMVTQTQAGNSRSDAMPQARQACLTGMDRQCEQVEE
 HAPRLQTVSSGMDAHMAMCALGTCATPFPFLQCIHSHP"
 565..801
 /gene="Sp-B"
 /product="pulmonary surfactant-associated protein B"
 BASE COUNT 326 a 565 c 450 g 319 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.11 Length: 1660
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.238 Indels: 0
 Gaps: 0
 DB: 4
 US-10-024-955-7 (1-213) x AF107544 (1-1660)
 QY 188 AaIlePhcInaSpHrValarglys 196
 LOCUS |||||||||||||||||||||||||
 DB 256 GCCATTTCACGACACAGCGCAAA 282
 RESULT 10
 AC017321 21667 bp DNA linear HTG 09-DEC-1999
 LOCUS AC017321
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 pieces.
 ACCESSION AC017321
 VERSION AC017321.1 GI:6553665
 KEYWORDS HTG; HTGS; PHASE2.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 TITLE 1 (bases 1 to 21667)
 AUTHORS Adams, M. and Venter, J.C.
 JOURNAL Direct Submission
 JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10210290 by the submitter.
 For more information on this record e-mail to flycelera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES Location/Qualifiers
 source 1..21667
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 6617 a 4534 c 4317 g 6199 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 60.7 Length: 21667
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.238 Indels: 0
 Gaps: 0
 DB: 2

US-10-024-955-7 (1-213) x AC017321 (1-21667)

QY 4 PheleuleuilealaialaAlaph 12
 |||||||||||||||||||

Db 10414 TTTTGTCAATGCTGCTGTTCTTT 10440

RESULT 11
 AP003544 135792 bp DNA linear HTG 21-MAR-2002
 LOCUS AP003544
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone P0599C12,
 *** SEQUENCING IN PROGRESS *** in ordered pieces.

ACCESSION AP003544
 VERSION AP003544.1 GI:13810550
 KEYWORDS HTG: HTGS_PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
 clone: P0599C12.

ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
 TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
 clone: P0599C12
 JOURNAL Published Only in Database (2001)
 REFERENCE 2 (bases 1 to 135792)
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
 JOURNAL Direct Submission
 Submitted (25-APR-2001) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program: Kannondai
 2-1-1, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@agr.affrc.go.jp, URL: http://rqp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..135792
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="6"
 /clone="P0599C12"

BASE COUNT 37908 a 30304 c 30206 g 37124 t 250 others

ORIGIN

Alignment Scores:
 Pred. No.: 313 Length: 135792
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.23% Indels: 0
 DB: 2 Gaps: 0

US-10-024-955-7 (1-213) x AP003544 (1-135792)

QY 181 ValleyserspyalleuThralaile 189
 |||||||||||||||||||

Db 4525 GTATTATCGACGTATTACGCCATT 4551

RESULT 12
 AC126526/c 146738 bp DNA linear HTG 13-JUL-2002
 LOCUS AC126526/c
 DEFINITION Rattus norvegicus clone CH230-240E2, *** SEQUENCING IN PROGRESS
 *** 37 unordered pieces.

ACCESSION AC126526
 VERSION AC126526.1 GI:21700493
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 146738)
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alshrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barberia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dedetich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichatage,O., Lien,C., Liu,J., Liu,W., Louised,H.,
 Lozado,J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapa,P., Martin,R., Meritindate,A., Martinez,E.,
 Massey,E., Mawliny,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Moser,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwoudu,G.,
 Otagunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojubokan,I., Rolte,N., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoshari,N., Slason,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalona,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,T.F., Zhou,J., Zorilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.

REFERENCE 2 (bases 1 to 146738)
 AUTHORS Worley,K.C.
 JOURNAL Direct Submission
 Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 146738)
 Worley,K.C.
 JOURNAL Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GMEL
 Center clone name: CH230-240E2
 Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 114654 bases at least Q40
 Consensus quality: 118931 bases at least Q30
 Consensus quality: 121615 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 37 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1007: contig of 1007 bp in length
 1008 1107: gap of unknown length
 1108 2806: contig of 1699 bp in length
 2807 2906: gap of unknown length
 2907 4375: contig of 1469 bp in length
 4376 4475: gap of unknown length
 4476 5556: contig of 1081 bp in length
 5557 5656: gap of unknown length
 5657 6714: contig of 1058 bp in length
 6715 8240: contig of 1426 bp in length
 8241 8341: gap of unknown length
 8341 9893: contig of 1533 bp in length
 9894 9994: gap of unknown length
 9994 11387: contig of 1394 bp in length
 11388 12640: contig of 1153 bp in length
 12641 12741: gap of unknown length
 12741 14664: contig of 1924 bp in length
 14665 14764: gap of unknown length
 14765 16896: contig of 2132 bp in length
 16897 16997: gap of unknown length
 16997 18581: contig of 1585 bp in length
 18582 18681: gap of unknown length
 18682 20010: contig of 1329 bp in length
 20011 20110: gap of unknown length
 20111 21276: contig of 1166 bp in length
 21277 21376: gap of unknown length
 21377 22714: contig of 1338 bp in length
 22715 22814: gap of unknown length
 22815 24723: contig of 1909 bp in length
 24724 24823: gap of unknown length
 24824 26030: contig of 1207 bp in length
 26031 26130: gap of unknown length
 26131 28853: contig of 2723 bp in length
 28854 28953: gap of unknown length
 28954 31364: contig of 2411 bp in length
 31365 31464: gap of unknown length
 31465 34609: contig of 3145 bp in length
 34610 34709: gap of unknown length
 34710 37928: contig of 3219 bp in length
 37929 38028: gap of unknown length
 38029 39742: contig of 1714 bp in length
 39743 39842: gap of unknown length
 39843 42683: contig of 2841 bp in length
 42684 42783: gap of unknown length
 42784 46156: contig of 3373 bp in length
 46157 46256: gap of unknown length
 46257 49615: contig of 3359 bp in length
 49616 49715: gap of unknown length
 49716 54761: contig of 5046 bp in length
 54762 54861: gap of unknown length
 54862 59705: contig of 4844 bp in length
 59706 59805: gap of unknown length
 59806 63369: contig of 3564 bp in length
 63370 63469: gap of unknown length
 63470 69441: contig of 5972 bp in length

* 69442 69541: gap of unknown length
 * 69542 76397: contig of 6856 bp in length
 * 76398 76497: gap of unknown length
 * 76498 84104: contig of 7607 bp in length
 * 84105 84204: gap of unknown length
 * 84205 91939: contig of 7735 bp in length
 * 91940 92039: gap of unknown length
 * 92040 100255: contig of 8216 bp in length
 * 100256 100355: gap of unknown length
 * 100356 110204: contig of 9849 bp in length
 * 110205 110304: gap of unknown length
 * 110305 123625: contig of 13321 bp in length
 * 123626 123725: gap of unknown length
 * 123726 134498: contig of 10773 bp in length
 * 134499 134598: gap of unknown length
 * 134599 146738: contig of 12140 bp in length.
 FEATURES
 source 1. 146738
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-240E2"
 BASE COUNT 36063 a 32001 c 32143 g 37660 t 8871 others
 ORIGIN

Alignment Scores:
 Pred. No.: 336 Length: 146738
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.23% Indels: 0
 DB: 2 Gaps: 0

US-10-024-955-7 (1-213) x AC126526 (1-146738)

OY 118 leuAlaTYrLysLeuGlyAspLeuHis 126
 Db 18970 CTACCTTAATACTTGAGATCTCAT 18944

RESULT 13

AC012070 149776 bp DNA linear PRI 14-JUL-2002
 LOCUS Homo sapiens BAC clone RP11-318K6 from 2, complete sequence.
 DEFINITION AC012070
 ACCESSION AC012070.8 GI:21747558
 VERSION
 KEYWORDS HGC.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 149776)
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792

REFERENCE
 2 (bases 1 to 149776)
 Isak, A. and Abbott, A.
 TITLE The sequence of Homo sapiens BAC clone RP11-318K6
 JOURNAL Unpublished (2001)

REFERENCE
 3 (bases 1 to 149776)
 Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-OCT-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
 4 (bases 1 to 149776)
 Waterston, R.H.

REFERENCE
 5 (bases 1 to 149776)
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
 6 (bases 1 to 149776)
 Waterston, R.
 TITLE Direct Submission

```

JOURNAL      Submitted (09-AUG-2001) Department of Genetics, Washington
REFERENCE    University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS      6 (bases 1 to 149776)
TITLE        Waterston,R.
JOURNAL      Direct Submission
REFERENCE    Submitted (07-NOV-2001) Department of Genetics, Washington
AUTHORS      University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE        7 (bases 1 to 149776)
JOURNAL      Waterston,R.
REFERENCE    Direct Submission
COMMENT      Submitted (14-JUL-2002) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Jul 14, 2002 this sequence version replaced g1:13162544.
              ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@wustl.wustl.edu
              ----- Summary Statistics
              Center project name: H_NH0318K06
              -----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPlCt-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RPl1-335M11, 200 bp overlap; the
clone sequenced to the right is RPl1-674L1. Actual start of this
clone is at base position 150598 of RPl1-335M11; actual end is at
base position 149776 of RPl1-318K6.

There are polymorphic base pair differences in the overlap between
the clone RPl1-318K6 and RPl1-674L1.

FEATURES
Source
Location/Qualifiers
1..149776
/oranism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RPl1-318K6"
/clone_lib="RPlCt-11"
420..484
/rpc_family="MER1_type"
485..799
/rpc_family="Alu"
844..962
/rpc_family="MER1_type"
repeat_region
repeat_region
repeat_region
repeat_region

```

repeat_region	2744.3104	/rpt_family="MALR"
repeat_region	3477.3499	/rpt_family="AT_rich"
repeat_region	3850.4119	/rpt_family="Alu"
repeat_region	4386.4511	/rpt_family="MIR"
repeat_region	4844.5098	/rpt_family="Mariner"
repeat_region	5226.5361	/rpt_family="L1"
repeat_region	5410.5452	/rpt_family="AT_rich"
repeat_region	6237.6317	/rpt_family="MIR"
repeat_region	6539.6858	/rpt_family="Alu"
repeat_region	7015.7042	/rpt_family="AT_rich"
repeat_region	7213.7234	/rpt_family="(CA)n"
repeat_region	7235.7258	/rpt_family="(GA)n"
repeat_region	8187.8491	/rpt_family="Alu"
repeat_region	15696.15812	/rpt_family="CR1"
repeat_region	16452.16174	/rpt_family="AT_rich"
repeat_region	16617.16647	/rpt_family="AT_rich"
repeat_region	16701.16729	/rpt_family="AT_rich"
repeat_region	18438.18727	/rpt_family="AT_rich"
repeat_region	19224.19249	/rpt_family="(TG)n"
repeat_region	19519.19818	/rpt_family="Alu"
repeat_region	19976.20298	/rpt_family="L1"
repeat_region	20523.20745	/rpt_family="L1"
repeat_region	20776.20938	/rpt_family="L2"
repeat_region	21146.21278	/rpt_family="L1"
repeat_region	22202.22349	/rpt_family="L1"
repeat_region	22471.22795	/rpt_family="L1"
repeat_region	22887.23259	/rpt_family="L1"
repeat_region	23877.24244	/rpt_family="ERVL"
repeat_region	24250.24291	/rpt_family="MALR"
repeat_region	25480.25507	/rpt_family="(TC)n"
repeat_region	26053.26084	/rpt_family="AT_rich"
repeat_region	26097.26149	/rpt_family="AT_rich"
repeat_region	26201.26278	/rpt_family="L1"
repeat_region	26329.26635	/rpt_family="ERV1"
repeat_region	26672.26962	/rpt_family="ERV1"
repeat_region	26990.27094	/rpt_family="Alu"
repeat_region	27310.27412	/rpt_family="ERV1"

repeat_region /rpt_family="Alu" 27541..27700
repeat_region /rpt_family="MIR" 28454..28469
repeat_region /rpt_family="AT-rich" 28963..29261
repeat_region /rpt_family="Alu" 29324..30139
repeat_region /rpt_family="L1" 30150..31521
repeat_region /rpt_family="L2" 31549..31616
repeat_region /rpt_family="GA-rich" 31889..32075
repeat_region /rpt_family="MIR" 32334..32371
repeat_region /rpt_family="AT-rich" 33317..33521
repeat_region /rpt_family="MIR" 34282..34329
repeat_region /rpt_family="L2"

Alignment Scores:
Pred. No.: 342 Length: 149776
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x AC012070 (1-149776)

OY 141 AlaleuSerLeuGlutIleSerAspGlu 149
Db 34837 GCTCTTCTCTGGAGATCTCTGATGAA 34863

RESULT 14
AP003344/c
LOCUS AP003344 155633 bp DNA linear PLN 10-AUG-2002
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
PAC clone:P0413G02.
ACCESSION AP003344
VERSION AP003344.2 GI:21715992
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:P0413G02.
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC
clone:P0413G02
Published only in Database (2001)
2 (bases 1 to 155633)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (28-FEB-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Jul 9, 2002 this sequence version replaced gi:13359080.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI Nonredundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologues of the coding regions were searched against
NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.

COMMENT

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.
The orientation of the sequence is from SP6 to T7 of the PAC clone. This sequence of P0413G02 clone has an overlap with P049D09 clone (DBJ : AP003265) at the position 1 to 13,040 of 5' end and an overlap with P0010B10 clone (DBJ: AP003224) at the position 155,178 to 155,633 of 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/genomeseq.html>.

FEATURES
source
location/Qualifiers
1..155633
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0413G02"
/complement(join(3362..3691,4426..4689,5176..5445,6412..6537,6690..7298))
/gene="P0413G02.1"
/complement(join(3362..3691,4426..4689,5176..5445,6412..6537,6690..7298))
/gene="P0413G02.1"
/codon_start=1
/product="putative cytokinin oxidase"
/protein_id="BAC07345.1"
/db_xref="GI:22202687"
/translation="MANCLVFWFLIYCLISVNGVLPVAPADPAAADGCGGRSLVEPSDWEASIDREGRLTSAERPLAFHRRGAGDAVALKAAAGSAGSIRVARSRGHSISGQAAGAGVVDMSHGMRAERLTPVTSFALGHTYIDWGGELMDVLAWMGLDAPRSWTIDLYLSVDSGQFLNSAGISGQAFHHPQISNVYELDVVTGKEVTCSSNNPDLFFGALGGLGLOLIIIRARIALEPAPRPMIRALYSNFEFPADQERLISQHGGRFDYEGFVVAEGLIINMRSSFFSPONPVKSLSKHSGVLCLEYTKANDSDTAFTVDVEALIGELINIPGTFPTTDLPYDPLDRVAKELKLRGKMGMEFVPLMNLFPVAPASRTADRGVFRGVLSRTNRGGLIILYPMNRHNDPPSSVYTPEDVFYVPLRLSVAPGSDPAPQSLAELEKRNREILFCDEAGIKOPLPHNKAOREMHAHFKAARFARLKAEEPPRAMLATGGGIFDSPPLAES"
join(24594..24698,25255..25542)
/gene="P0413G02.2"
join(24594..24698,25255..25542)
/gene="P0413G02.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC07346.1"
/db_xref="GI:22202688"
/translation="MDCRPSQPDYSPFASPKMTSGRGEKTYQSCIRIOYGLKNACL VFTSKHSCSNRPSTSNVMLHACIAFDVTSERKNNARSSSHSDSTSETSMANVRRLNPKHNDIELVRLPLVILFDYTESLV"
complement(join(25838..26149,26353..26568,26706..27047,27183..27425,27535..27987,28303..28416,29765..29836,30157..30195,30432..30523,30627..30738,31226..31321,31408..31482,31561..31704,32601..32741,33920..34106,34561..34648,34973..35024))
/gene="P0413G02.3"
complement(join(25838..26149,26353..26568,26706..27047,27183..27425,27535..27987,28303..28416,29765..29836,30157..30195,30432..30523,30627..30738,31226..31321,31408..31482,31561..31704,32601..32741,33920..34106,34561..34648,34973..35024))
/gene="P0413G02.3"
/note="contains EST D48148(S14217) similar to Arabidopsis thaliana chromosome 1, F6F9_23 unknown protein"
/codon_start=1
/protein_id="BAC07347.1"
/db_xref="GI:22202689"
/translation="MAAVELGACHKNDGRVERDLNTVAESMKHEALFEKELPLPSADIVNKKVKKIPFMSASPAAGRGTASLSPSSCCPPFPFTHSPRISASPAAMDLDPLRYENVSVDNDVANIYLSYLMHNCFRFETAFETFLSTGLLELPDYTVDVDRKRKFESF

VLGNALKAIDTEELAPNLENDMDLHFDLSLHFIELVRSRKTALBEGOKKLTFE
GKVPKVERKERTNDVKEHLHYLLPIORKNRNIGLEKTFESMDFMALLAYEPEKSP
MFLHSPYRONVADSLNRAVLNANRPASLERVIOQTVOYXLOOVGQSDYD
FSLKAFNLNIVSLNOSAVHLLCLMGLKVLIVITGGAIDISIKIEYDOKSWSE
KKGDCGCTDOVKLDYPOEILITVSGYGSJGCGIVNSILTRSNLSTKYCPGSGEG
TPPSLVAATGKVIYFGHKSQFPLDSTIGCFHKEKATPSNAPSALRSTTRHDNG
NRADSNAGIDWVLAVRDGDYSVLTSNPKQYPNQSDATLWKNWLSLPSFYDN
GTMITSTPYAFGFWGNGGTIFDDGIYGVQINLRGLGISTMKVLYDRNGAIGD
KRGSSGAAREKVPFDEPSEILITHITGFESTMIGSTVKSILFTTKSHGPGDE
TCTFESSCLTEGRIYVGHGDMWIDISIGVHVEGLSORADALLETSPSRHML
AVANORIGDEVYGVKPEIPYNGCPWGGGCKPMDGVYTGKQIYIMADITGSVQ
LEYDRSGQSTWSTRHNGGQITRIKIDIPHVELNCTIGYNTCODBGPVLSTILV
SNRGKYPGPEEYGTFFSATTEGKVVGFHGRGLYLDAIGVHMOMLGDNRNTAAPS
SNKYISKLYF
36888..39171
/gene="P0413602.4"
36888..39171
/gene="P0413602.4"
/note="probably inactive due to stop codon(s) in CDS
pseudogene, similar to Oryza sativa hypothetical protein,
P0439E07.4"
/pseudo
/join(40751..40897,41400..41807)
/gene="P0413602.5"
/join(40751..40897,41400..41807)
/gene="P0413602.5"
/note="hypothetical protein
similar to Oryza sativa chromosome 1, P0499C11.2"
/codon_start=1
/protein_id="BAC07348.1"
/db_xref="GI:22202690"
/translation="MQINETIDKEFFTEVRKLAGLIPVNASPTPLNHEATARLLC
OGVKRTSRRAIIEPTOPGRSDHSVAGQTVGSRPSPARSDMLRLMCLATKQ
TCGAGOTGLRPVYIPOFTOLNRLVLRKRYLLRVSEFVGLGYKLIOAIDPMNPN
SKHKRSGCRATPKDPDITP"
/join(43411..43569,44090..44266)
/gene="P0413602.6"
/join(43411..43569,44090..44266)
/gene="P0413602.6"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC07349.1"
/db_xref="GI:22202691"
/translation="MLEGENASSISAQASERRPVRERGSGFVFAWMTDNTPEF
RCHFEQMCNPNIEFTIGFVNPDAEQLVGTGLAIEVVFALLVHMVATRRHNTSLTV
FSPSLCORV"
45510..46481
/gene="P0413602.7"
45510..46481
/gene="P0413602.7"
/note="hypothetical protein
similar to Oryza sativa chromosome 1, P0490D09.14"
/codon_start=1
/protein_id="BAC07350.1"
/db_xref="GI:22202692"
/translation="MALRLSLLLVAQALLVVVGASATVARGTOLYTTGYPOQID
SPAPMATSTGSCNCTAFRANLLTLASLPDQAAPTGASMGAGAGGAGGDDDDFA
RSGCLDSPTSCQRCICLAVIDVAGCGADTTRAGAMLSGCYLAADPTATSPRESA
FHRWFDGDTLPFSDNIDPTFLDLAGVSAVLAALAAANRSGRTMLAAOEDDGN
TGLAGVLAQACAGRAPADCVQCLDSARAMPCCNNAMNIGSVAVALSIDCVLOYH
HAPGLSGGKRWHRSSPSVCPRRWSSASNNHVVVVVVESSSDNIO"
51733..52581
/gene="P0413602.8"
51733..52581
/gene="P0413602.8"
/note="contains ESTs AU056020(S20227), AU056021(S20227)
similar to Oryza sativa chromosome 1, P0490D09.14
unknown protein"
/codon_start=1
/protein_id="BAC07351.1"
/db_xref="GI:22202693"
/translation="MAITKLSCLLCVYSTANGARYFPQIDCSPAFTSSNGTAFR
ANLLAIPTDLPDPAQAAATRPASTQAGGDRALLSVCLGSDTALYRESIAAVADV

gene
CDS
Alignment Scores:
Pred. No.: 354 Length: 155633
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 8 Gaps: 0
US-10-024-955-7 (1-213) x AP003344 (1-155633)
QY 181 ValLeuSerAspValLeuThrAlaIle 189
Db 79701 GTATTATCGACGATTAACGGCCATT 79675
RESULT 15
AP004731/c 158456 bp DNA linear HTG 21-MAR-2002
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) chromosome 6 clone
OSJNBa0016D02, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION
AP004731
VERSION
AP004731.1 GI:18656377
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:OSJNBa0016D02.
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriopharidaceae; Oryzaceae; Oryza.
REFERENCE
1 Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, BAC
clone:OSJNBa0016D02
Published Only in Database (2002)
2 (bases 1 to 158456)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (13-FEB-2002) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,
tel:81-298-38-7441, fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
1..158456
FEATURES
source

```

/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="6"
/clone="OSNBS0016D02"

BASE COUNT    44183 a 35607 c 35786 g 42641 t    239 others
ORIGIN

Alignment Scores:
Pred. No.:      360      Length:      158456
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      4.23%      Indels:      0
DB:                2      Gaps:      0

US-10-024-955-7 (1-213) x AP004731 (1-158456)

QY    181  ValLeuSerAspValLeuThrAlaIle 189
      |||||
Db    97255  GATATATCGAGCATATTAACGCCCAT 97229

RESULT 16
AC009345      158895 bp  DNA  linear  INV 22-MAR-2001
LOCUS        Drosophila melanogaster, chromosome 3R, region 98C-98C, BAC clone
AC009345
VERSION      AC009345.7  GI:13430982
KEYWORDS     HTG.
SOURCE       Drosophila melanogaster.
ORGANISM     Drosophila melanogaster.
              Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 158895)
AUTHORS      Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
              Holt,R.A., Evans,C.A., Gocayne,J.D., Amentlides,P.G., Brandon,R.C.,
              Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
              Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
              Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
              Ferreira,S., Frise,E., Galle,R.F., Gard,N.S., George,R.A.,
              Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
              Idegam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
              McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
              Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
              Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
              Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M.,
              Zaveri,J.S., Smith,H.O., Rubin,G.M. and Ventler,J.C.
              Sequencing of Drosophila chromosome 3R, region 98C-98C
              Unpublished
              2 (bases 1 to 158895)
REFERENCE    Celisner,S.E., Agbayan,I.A., Arcalina,T.T., Baxter,E., Blazej,R.G.,
              Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
              Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
              Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
              Kim,E., Lee,B., Lewis,S., Li,P., Lomontan,M.A., Mazda,P.,
              Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,J.M., Park,S.,
              Pfeiffer,B., Poon,H., Sequiera,A., Sethi,H., Snil,E.,
              Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
              Rubin,G.M.
              Direct Submission
              Submitted (17-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
              Laboratory, MS 64-121, Berkeley, CA 94720, USA
              On Mar 22, 2001 this sequence version replaced gi:6996980.
              Sequence submitted by:
              Lawrence Berkeley National Genome Project
              Berkeley Berkeley National Laboratory, MS 64-121
              Berkeley, CA 94720
              This sequence was assembled using end sequences from a whole genome
              shotgun and from subclones of this BAC and its neighboring clones.
              For further information about this sequence, including its location
              and relationship to other sequences, please visit our sequence

```

```

archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1. 158895
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="98C-98C"
/clone="BACR28F20 (D1064)"
/clone_id="RPC1-98 (Roswell) Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
PBACE3.6)"

BASE COUNT    45558 a 34576 c 34201 g 44560 t
ORIGIN

Alignment Scores:
Pred. No.:      361      Length:      158895
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      4.23%      Indels:      0
DB:                3      Gaps:      0

US-10-024-955-7 (1-213) x AC009345 (1-158895)

QY    4  PheLeuLeuIleAlaValAlaPhe 12
      |||||
Db    57243  TTTTGCTCATTCGCTGCTGCTTTT 57269

RESULT 17
AC073336/c    166261 bp  DNA  linear  PRI 07-NOV-2001
LOCUS        Homo sapiens BAC clone RP11-476H24 from 7, complete sequence.
AC073336
VERSION      AC073336.6  GI:13194991
KEYWORDS     HTG.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens.
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 166261)
REFERENCE    Toward a complete human genome sequence
              Sulston,J.E. and Waterston,R.
              Genome Res. 8 (11), 1097-1108 (1998)
              9847074
MEDLINE     99063792
PUBMED      9847074
REFERENCE    2 (bases 1 to 166261)
AUTHORS      Belter,E., Abbott,A. and Burkett,M.
              The sequence of Homo sapiens BAC clone RP11-476H24
              Unpublished
              3 (bases 1 to 166261)
REFERENCE    Waterston,R.H.
              Direct Submission
              Submitted (14-JUN-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              4 (bases 1 to 166261)
REFERENCE    Waterston,R.H.
              Direct Submission
              Submitted (04-MAR-2001) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              5 (bases 1 to 166261)
REFERENCE    Waterston,R.
              Direct Submission
              Submitted (09-MAY-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              6 (bases 1 to 166261)
REFERENCE    Waterston,R.
              Direct Submission
              Submitted (10-MAY-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

REFERENCE 7 (bases 1 to 166261)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Mar 4, 2001 this sequence version replaced gi:11415220.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: MUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0476H24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPheerson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GEN/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tateono,M., Caranese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-419K23, 200 bp overlap; the clone sequenced to the right is RP11-5C23. Actual start of this clone is at base position 126203 of RP11-419K23; actual end is at base position 166261 of RP11-476H24.

RP11-476H24 contains a tandem repeat between bases 19178 to 19980 where the sequence fidelity can not be guaranteed. Digest information suggests that there are 600 bps missing from the current assembly.

FEATURES

Source 1..166261
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"
 /clone="RP11-476H24"
 /clone_lib="RPCT-11"
 /clone_1lb="373..460"
 repeat_region /rpt_family="ERV1"
 repeat_region 461..740
 /rpt_family="Alu"
 repeat_region 741..867
 /rpt_family="ERV1"
 repeat_region 1442..1739

misc_feature	/rpt_family="Alu"	1444..1969	/note="similar to EST AL354867 (MID:g9797187)"
repeat_region	/rpt_family="MALR"	1863..2062	
repeat_region	/rpt_family="MALR"	2063..2324	
repeat_region	/rpt_family="MALR"	2325..2612	
repeat_region	/rpt_family="L1"	2613..2794	
repeat_region	/rpt_family="MALR"	2797..2855	
repeat_region	/rpt_family="MALR"	3655..4024	
misc_feature	/rpt_family="Alu"	4757..5099	/note="similar to EST AW894193 (MID:g8058398)"
repeat_region	/rpt_family="MALR"	4965..5217	
misc_feature	/rpt_family="MALR"	5192..5593	
repeat_region	/note="similar to EST AW895143 (MID:g8059348)"	5513..5691	
repeat_region	/rpt_family="MALR"	5692..5806	
repeat_region	/rpt_family="L1"	6916..7194	
misc_feature	/rpt_family="Alu"	7231..7379	/note="similar to EST AL354867 (MID:g9797187)"
misc_feature	/note="similar to EST AL354867 (MID:g9797187)"	8136..8265	
misc_feature	/note="similar to EST AL354867 (MID:g9797187)"	10020..10176	
repeat_region	/note="similar to EST AI370936 (MID:g4149689) ta06h03.x1"	10588..10729	
repeat_region	/rpt_family="MALR"	10730..11069	
repeat_region	/rpt_family="MALR"	11070..11296	
repeat_region	/rpt_family="MALR"	11833..12300	
repeat_region	/rpt_family="MALR"	13450..13547	
misc_feature	/note="similar to EST AI370936 (MID:g4149689) ta06h03.x1"	15546..15732	
repeat_region	/rpt_family="MERL_type"	15744..15918	
repeat_region	/rpt_family="MIR"	15948..16073	
repeat_region	/rpt_family="L2"	16077..16244	
repeat_region	/rpt_family="ERV1"	16370..16523	
repeat_region	/rpt_family="L2"	16642..16800	
repeat_region	/rpt_family="L2"	16805..16892	
repeat_region	/rpt_family="(TAGA)n"	19244..19420	
repeat_region	/rpt_family="(CCA)n"	19444..19969	
repeat_region	/rpt_family="(CCA)n"	20090..20175	
repeat_region	/rpt_family="MIR"	20311..20371	
repeat_region	/rpt_family="MERL_type"	21427..21479	
repeat_region	/rpt_family="(TGGG)n"	21533..21705	
repeat_region	/rpt_family="(TGCA)n"	21693..21832	
repeat_region	/rpt_family="(TAGA)n"	23253..23378	
repeat_region	/rpt_family="MERL_type"		


```
repeat_region 23379..23660
                /rpt_family="Alu"
repeat_region 23661..23719
                /rpt_family="MERL_type"
misc_feature 24318..24651
                /note="similar to EST T02925 (NID:9314166)"
repeat_region 25770..26060
                /rpt_family="L1"
repeat_region 26130..26310
                /rpt_family="L1"
repeat_region 26352..26953
                /rpt_family="L1"
repeat_region 26954..27318
```

Alignment Scores:

Pred. No.:	376	Length:	166261
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.23%	Indels:	0
DB:	9	Gaps:	0

US-10-024-955-7 (1-213) x AC073336 (1-166261)

Qy 179 PheglyValLeuSerAspValLeuThr 187
Db 12932 TTCGGGGTGTGTCTGATCTCTCACT 12906

RESULT 18
AC013177/c
LOCUS AC013177 167596 bp DNA linear HTG 03-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION AC013177
VERSION AC013177.1 GI:6223154
KEYWORDS HTG: HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 167596)
REFERENCE Adams,M. and Venter,J.C.
AUTHORS Direct Submission
JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDW:10213925 by the submitter.
For further information on this sequence you may e-mail to
fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..167596
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 47994 a 36373 c 3566 g 47363 t
ORIGIN

Alignment Scores:

Pred. No.:	378	Length:	167596
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.23%	Indels:	0
DB:	2	Gaps:	0

US-10-024-955-7 (1-213) x AC013177 (1-167596)

Qy 8 AlaAlaValAlaPheValAlaValSer 16
Db 149722 GCACGAGTCGATTTGGCTGTCTGCT 14966

RESULT 19
AC117360
LOCUS AC117360
DEFINITION Rattus norvegicus clone CH230-396E2, *** SEQUENCING IN PROGRESS
ACCESSION AC117360
VERSION AC117360.2 GI:21671557
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 173476)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbarka,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Mattindale,A., Martinez,E.,
Massey,E., McWhiney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguni,M., Okunolu,G.,
Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Roife,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleciyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 173476)
Worley,K.C.
Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 173476)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 2, 2002 this sequence version replaced gi:20127297.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GUSX
 Center clone name: CH230-396E2
 ----- Summary Statistics -----
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 144345 bases at least Q40
 Consensus quality: 150228 bases at least Q30
 Consensus quality: 155431 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 39 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 1013: contig of 1013 bp in length
 * 1014 1113: gap of unknown length
 * 1114 2312: contig of 1199 bp in length
 * 2313 2412: gap of unknown length
 * 2413 4197: contig of 1785 bp in length
 * 4198 4297: gap of unknown length
 * 4298 6287: contig of 2000 bp in length
 * 6298 6397: gap of unknown length
 * 6398 8112: contig of 1715 bp in length
 * 8113 8212: gap of unknown length
 * 8213 9256: contig of 1044 bp in length
 * 9257 9356: gap of unknown length
 * 9357 10763: contig of 1407 bp in length
 * 10764 10863: gap of unknown length
 * 10864 12892: contig of 2029 bp in length
 * 12893 12992: gap of unknown length
 * 12993 14574: contig of 1582 bp in length
 * 14575 14674: gap of unknown length
 * 14675 16493: contig of 1819 bp in length
 * 16494 16593: gap of unknown length
 * 16594 18681: contig of 2088 bp in length
 * 18682 18781: gap of unknown length
 * 18782 20325: contig of 1544 bp in length
 * 20326 20425: gap of unknown length
 * 20426 23648: contig of 3223 bp in length
 * 23649 23748: gap of unknown length
 * 23749 25502: contig of 1754 bp in length
 * 25503 25602: gap of unknown length
 * 25603 27957: contig of 2355 bp in length
 * 27958 28057: gap of unknown length
 * 28058 30970: contig of 2913 bp in length
 * 30971 31070: gap of unknown length
 * 31071 33208: contig of 2138 bp in length
 * 33209 33308: gap of unknown length
 * 33309 36683: contig of 3375 bp in length
 * 36684 36783: gap of unknown length
 * 36784 38950: contig of 2167 bp in length
 * 38951 39050: gap of unknown length
 * 39051 41148: contig of 2098 bp in length
 * 41149 41248: gap of unknown length
 * 41250 44753: contig of 3505 bp in length
 * 44754 44853: gap of unknown length
 * 44854 47871: contig of 3018 bp in length
 * 47872 47971: gap of unknown length
 * 47972 51288: contig of 3317 bp in length
 * 51289 51388: gap of unknown length
 * 51389 54260: contig of 2872 bp in length
 * 54261 54360: gap of unknown length
 * 54361 58331: contig of 3771 bp in length
 * 58332 58431: gap of unknown length

58432 62672: contig of 4241 bp in length
 * 62673 62772: gap of unknown length
 * 62773 66712: contig of 3940 bp in length
 * 66713 66812: gap of unknown length
 * 66813 73058: contig of 6246 bp in length
 * 73059 73158: gap of unknown length
 * 73159 79228: contig of 6070 bp in length
 * 79229 79328: gap of unknown length
 * 79328 84562: contig of 5234 bp in length
 * 84563 84662: gap of unknown length
 * 84663 91677: contig of 7015 bp in length
 * 91678 91777: gap of unknown length
 * 91778 99397: contig of 7620 bp in length
 * 99398 99497: gap of unknown length
 * 99498 109651: contig of 10154 bp in length
 * 109652 109751: gap of unknown length
 * 109752 118288: contig of 8537 bp in length
 * 118289 118388: gap of unknown length
 * 118389 127339: contig of 8851 bp in length
 * 127340 127339: gap of unknown length
 * 127340 136960: contig of 9621 bp in length
 * 136961 137060: gap of unknown length
 * 137061 146354: contig of 9294 bp in length
 * 146355 146454: gap of unknown length
 * 146455 160290: contig of 13836 bp in length
 * 160291 160390: gap of unknown length
 * 160391 173476: contig of 13086 bp in length.
 FEATURES
 source Location/Qualifiers
 1..173476
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-396E2"
 BASE COUNT 49031 a 35836 c 34403 g 49020 t 5186 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 390 Length: 173476
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.23% Indels: 0
 DB: 2 Gaps: 0
 US-10-024-955-7 (1-213) x AC117360 (1-173476)
 QY 170 GlycylleuSerIleuAspProIle 178
 |||||||
 Db 105459 GGAGGCGTTCATTCATGATCCATC 105485
 RESULT 20
 AC007257/c
 LOCUS AC007257
 DEFINITION Drosophila melanogaster, chromosome 2L, region 30B-, BAC clone
 AC007257
 ACCESSION AC007257
 VERSION AC007257.9 GI:5685883
 KEYWORDS HTG.
 ORGANISM Drosophila melanogaster.
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 183007)
 REFERENCE
 AUTHORS
 Celniker, S.E., Aabayani, A., Arcalata, T.T., Baxter, E., Blazek, R.G.,
 Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hinkle, A., Hoskins, R.A., Houston, K.A., Humastli, S.R., Karra, K.,
 Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Ling, H.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,
 Pfeiffer, B., Poen, L., Richards, S., Seguela, A., Sethi, H.,
 Svitskas, R.R., Wan, K.H., Webster, D., Woolley, F., Yang, S., Yee, M.,
 Zierman, L.L., and Rubin, G.M.
 Sequencing of Drosophila chromosome 2L, region 30B-

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 183007)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazek,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shit,E., Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

TITLE Direct Submission
JOURNAL Submitted (08-APR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Aug 3, 1999 this sequence version replaced gi:5579492.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES
SOURCE location/Qualifiers
1. 183007
/organism="Drosophila melanogaster"
/strain="y: cn bw^{sp}"
/db_xref="taxon:7227"
/chromosome="2L"
/map="30B-"
/clone="BAC BACR24A22 (D574)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"
BASE COUNT 51847 a 39557 c 39297 g 52306 t
ORIGIN

Alignment Scores:
Pred. No.: 409 Length: 183007
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 3 Gaps: 0

US-10-024-955-7 (1-213) x AC007257 (1-183007)
QY 8 AlaAlaValAlaPheValAlaValSer 16
DB 135274 GCAGCAGTCGCAATTGTGCTGTCT 135248

RESULT 21
AC007819 189649 bp DNA linear INV 22-MAR-2001
LOCUS Drosophila melanogaster, chromosome 3R, region 98C-98C, BAC clone
DEFINITION BACR06K21, complete sequence.
ACCESSION AC007819
VERSION AC007819.8 GI:13430985
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 189649)
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amentides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreira,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,

McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanewong,S., Pittman,G.S., Puri,V., Richards,S., Schoefer,F., Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M., Zaveril,J.S., Smith,B.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 98C-98C
Unpublished
2 (bases 1 to 189649)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazek,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shit,E., Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

TITLE Direct Submission
JOURNAL Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Mar 22, 2001 this sequence version replaced gi:7230784.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES
SOURCE location/Qualifiers
1. 189649
/organism="Drosophila melanogaster"
/strain="y: cn bw^{sp}"
/db_xref="taxon:7227"
/chromosome="3R"
/map="98C-98C"
/clone="BACR06K21 (D787)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"
BASE COUNT 56401 a 40464 c 40247 g 52537 t
ORIGIN

Alignment Scores:
Pred. No.: 422 Length: 189649
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 3 Gaps: 0

US-10-024-955-7 (1-213) x AC007819 (1-189649)
QY 4 PheLeuIleAlaValAlaPhe 12
DB 182055 TTTTGTGCTATTCGTCGTCTTTT 182081

RESULT 22
AE003764 231732 bp DNA linear INV 05-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386035 section 89
DEFINITION OE 105, complete sequence.
ACCESSION AE003764 AE002708
VERSION AE003764.2 GI:10726819
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 231732)

SOICAGFONGNICETETGSSLVKQIHYSGLMTWLTIGIOSYGVSERRPDRTKGRKRGPG
ELEAATPANCVLYSHASDSROLRDGDCGYNSSFPDLHORCLK"

gene
 <193133. .>194351
 /gene="Mst98Ca"
 /note="CG11719"
 /map="98C2-98C3"
 /db_xref="FLYBASE:FBan0011719"
 /db_xref="FLYBASE:FBgn0002865"
 193133. .194351
 /gene="Mst98Ca"
 /product="CT5056"
 /note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."
 /db_xref="FLYBASE:FBan0011719"
 /db_xref="FLYBASE:FBgn0002865"
 193154. .194158
 /gene="Mst98Ca"
 /note="Mst98Ca gene product; Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"

CDS
 /protein_id="AAF56779.1"
 /db_xref="GI:7301663"
 /db_xref="FLYBASE:FBan0011719"
 /db_xref="FLYBASE:FBgn0002865"
 /translation="MCSPGCGPCDPCGCGPECSPKCYNAALBALPQCAPRIIPPF
 PCITVOOPRMICKRKYEMTEKIVPEPMVNRORITPRKYVDATRYTKVRLIWS
 OMVREPRYYYSMTDPDYVYVYPRKRCERREVCOSIICQRPQTIDIPPREYCYTP
 NPPIYKPSACPPCPIGFCAGGPGCCIPCHSTMOYPAESRCGPGCGGCGPGCG
 PRGCGPQPCGPGCGPGCAVPNCGGLTMPGPFVAPACPGCAPCGPGPL
 GNSPCGPGCGPCSPPCPYESPCEGPGCPAPFTWNTFHCGFVPGCGPQVPCGPGSPC
 "

gene
 <194511. .>195562
 /gene="Mst98Cb"
 /note="CG18396"
 /map="98C2-98C3"
 /db_xref="FLYBASE:FBan0018396"
 /db_xref="FLYBASE:FBgn0004171"
 194511. .195562
 /gene="Mst98Cb"
 /product="CT41800"
 /note="Nucleotide sequence of the Celera sequence differs

mRNA

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
505	9.00	231732	9	0	0	0	0
Percent Similarity:	100.00%						
Best Local Similarity:	100.00%						
Query Match:	4.23%						

DB: 3

US-10-024-955-7 (1-213) x AE003764 (1-231732)

Oy 4 PhleuleuleleAlaAlaValaIaphe 12
 |||||
 Db 167809 TTTTTCCTCATTCGCTGCTGCTTTT 167835

RESULT 23

AE003625
 LOCUS AE003625
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386055 section 18 of 63, complete sequence.
 ACCESSION AE003625
 VERSION AE003625.1
 KEYWORDS HTG.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydrioidae; Drosophilidae; Drosophila.
 1 (bases 1 to 260624)
 Adams,M.D., Celinker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,

George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Mortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Barton,R.C., Rogers,Y.H., Blazef,J.R., Champagne,O., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Heil,G., Nelson,C.R., Gabor,Miklos,G.L., Abriil,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotter,P., Butts,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,J., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Mishina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacile,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Rehnert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kimios,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R.P., Sun,E., Svateks,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Weissman,D.A., Welterstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,Q., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)

JOURNAL
 MEDLINE
 PUBMED
 10731132
 2 (bases 1 to 260624)
 Adams,M.D., Celinker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

FEATURES
 source
 gene
 mRNA
 CDS

Location/Qualifiers
 1. .260624
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2L"
 <1474. .>14937
 /gene="numb"
 /note="CG3779"
 /map="30B5-30B7"
 /db_xref="FLYBASE:FBan0003779"
 /db_xref="FLYBASE:FBgn0002973"
 join(11932. .2025,11691. .12012,12149. .12397,12582. .14937)
 /gene="numb"
 /product="CT12616"
 /note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."
 /db_xref="FLYBASE:FBan0003779"
 /db_xref="FLYBASE:FBgn0002973"
 join(11932. .12012,12149. .12397,12582. .13922)
 /gene="numb"
 /note="numb gene product; Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"
 /codon_start=1
 /protein_id="AAF52776.1"
 /db_xref="GI:7297521"
 /db_xref="FLYBASE:FBan0003779"
 /db_xref="FLYBASE:FBgn0002973"

```

/translaton="MGNSSSHTEPLEERGFTTRGKFGDVKNKGKSASFREKSPKKNDR
LRRSFRDSEFRRRKDRPESKPHQOMADEAVRSATCSFVKYLCGVEAFESGKMOYC
EALIKVLROSRDRVRGLHVSDDGLRVYDDETRGLIVDOTIEKVSQCADRKHGEF
SYICRBDGTRRRMGHGFACKDSGERLSHAGCAPAVCEKROKROKCEGVMTPTPK
NSFTPTGSSFRROOULTEBLAMATVGTNERSVDPCGSMRPPRAATYKPRPFAIERH
ATPMLEROSSEFRSTIGOSPFKROMSLINDPNSADRORAFLTRAAANPQOTPLR
VSPIAEVSPAKSAGADPLSAAVAADSVQLQELSQSLTLOTDALAAEDINE
NNRSINONITAAEQOVHVSYPAPTAQVTPVASTPTPYQTLHOSQSPSOSIET
STELPNAEOMLGHVRSSTSPAKRPTLYANVGRAGTILASGTAAGGCGGDDPFDE
MVANVAAKOLSDPLIPSTARSPLARHSNPTISPPKAPAGSFOYQL"
complement(<43607..>46415)
/gene="Gd1"
/note="CG4422"
/map="3088-3089"
/db_xref="FLYBASE:FBgn0004422"
/db_xref="FLYBASE:FBgn0004868"
complement(join(43607..44234,44317..44792,45911..46415))
/gene="Gd1"
/product="CT14400"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/db_xref="FLYBASE:FBan0004422"
/db_xref="FLYBASE:FBgn0004868"
complement(join(43773..44234,44317..44792,45911..46304))
/gene="Gd1"
/note="Gd1 gene product; Nucleotide sequence for this
sequence differs from the published sequence for this
transcript"
/codon_start=1
/protein_id="AAE52777.1"
/db_xref="GI:7297522"
/db_xref="FLYBASE:FBan0004422"
/db_xref="FLYBASE:FBgn0004868"
/translaton="MNEEYDAIVLGTGKCEILSGMLSVSGKKVLHIDRNKKYGGESA
STPLELRQVRGLPEPPGERGSRDMNVDLIRKFLMANGLIKLHIGVTVLEK
STEGSYVYKGRGIYAKVPVDOKEALASDLDMGEKRRRNRLTYVDREDDPTKWD
DPTKAMNGGLYDKFGLDKNTQDTGHALALFRDEYLINEPAVNTIRIKIYLSLAW
GKSPYLYPMYGLDELPOGARLSAIYGYTMDKPIDELVIGGGKVGVRSGSEYAK
CKOVYCDPSYVPELVRKRGKVICICILDHVASTDGLSTQIILPOKQVRSIDIY
SLVSTHOVAKMGFMVGWSTVETENPEVEIKRGLDLEIPAKQFVTIDYLRPIDD
GSSOIFISESYATHTFETTCLDVINIFRGRGTEPFDESKINHELGDEQG"
<55481..>56506
/gene="CG13112"
/evidence=not_experimental
/map="3089-3089"
/db_xref="FLYBASE:FBan0013112"
/db_xref="FLYBASE:FBgn0032120"
<55481..>56506
/gene="CG13112"
/product="CT32349"
/evidence=not_experimental
/db_xref="FLYBASE:FBan0013112"
/db_xref="FLYBASE:FBgn0032120"
55481..56506
/gene="CG13112"
/note="CG13112 gene product"
/evidence=not_experimental
/codon_start=1
/protein_id="AAE52778.1"
/db_xref="GI:7297523"
/db_xref="FLYBASE:FBan0013112"
/db_xref="FLYBASE:FBgn0032120"
/translaton="MPSANTEDLRKFLVVOQOARSAPNLTGTLSSGGINASASAT
TSLMTTSTGIGVANGAGEPERTVYGGNIPVGSAPGGGGGIPVMAAASRGHA
RSTSHGGATVANGGRPIKSGAMGHORAPGCGITDPSPSAPAPAGGSHRVSCKDF
ILPGRKREPAREPSAPTSKTCGRGSHROASRSSEYTLTLRTAPPMWKLJCNITF
GDKFEERSYTVVNHITVPKTRPDHPNQEGNKKIRTKYVLLSLIPNLLLEQFIR
VANLIFIFIVLMMWPEISAFGEVAMIPVLFLVGYAVADLEDRRRRASDRINNPT
TCRVYDG"
<66388..>71660
/gene="CG18419"
/map="3089-30810"
/db_xref="FLYBASE:FBan0018419"

```

```

mRNA
/db_xref="FLYBASE:FBgn0032121"
join(<66388..66452,67003..67183,67254..67433,67501..67625,
67684..68751,68819..69150,69220..69592,69661..69848,
69908..70247,70312..70497,70565..70765,70836..71023,
71083..71165,71407..71660)
/gene="CG18419"
/product="CT12667"
/db_xref="FLYBASE:FBan0018419"
/db_xref="FLYBASE:FBgn0032121"
join(66388..66452,67003..67183,67254..67433,67501..67625,
67684..68751,68819..69150,69220..69592,69661..69848,
69908..70247,70312..70497,70565..70765,70836..71023,
71083..71165,71407..71429)
/gene="CG18419"
/note="CG18419 gene product"
/codon_start=1
/protein_id="AAE52779.2"
/db_xref="GI:10728688"
/db_xref="FLYBASE:FBan0018419"
/db_xref="FLYBASE:FBgn0032121"
/translaton="MYKEKTPATRRRLIYIFPLSRETERHYKKVKKOELRVGDIVHS
NNEVPAADILLITSDPOGVCYIDTODLDGEMTLKRREVRGFEEMOSIFVPSKFSR
VEADAPTTRKIRRHGALLHPGGRVYISTECILLRSLKNDYITGIVYAHHEKRS
MLNNSGPRKRSQVEOQMNIDVICVITLITLCVGAIGCRMLLSSFTHPVPYLPN
KLRANMESMWIEWTYIVILQVMIPLSLVYITELCKILQVFIHNNVDLDAETNKQTE
GRAMNITEELGQIOHIFETDKTGTLENKMFRCVNGSDYNNHPELEKISKPGAP
APPLIPNDNISDMOULTOGTYITPAHORIOELVYLAIQNTYIVGAAAPHRDMNAG
LIEVOOIGNSPANKHKGOKOKLASTSTTTTITINGPTTPOVYSIPADRYTILA
ESRSVTPSPRNILPALPAOSHOPPLTSPISSASESPNSSESPPPKKRSLSNHS
PTGRKAVINLSKISTATFLNAKTQGRMKLPESKGTITRTDGRPLLEASPSDELA
LVNAAVSYDCLLNRSPNOILVSPMAGAREVILKVPEDSSRCMSLIVAQISO
EIVLYKRGADSSIMPLVLPCHSPNSPGLLEBOYOOLDRABRGLILVMAKRTLSA
DYIDWMAHROEIMESLENREERLRDSFAKLEMLTLGATGIDRLDGVPTIASL
SAGISVWVLGDKPEPAINAYSAKLFTOOMELIRLARSADAETAINPLYTDMND
KTSYTYGQGSLEKORALVYDOKTFTILDPKSKLILPRLRSKCAVILCCRSPL
OKAYLYVKKYEEINLRTLAGDANDVYSIOMADVGVISGDCGMAVMAADPTLRF
RYERLRLHAGVYCYORLSKMLLYFFRYKNAEFLFWOLYGFSGVQWMDMYLML
VNIIFLPLPLAAGSDVGIWERTGTITASCFLANVGRGVAIRPHADMLILLALYSL
VIFVFLCAVAESDVGIVMERGTITASCFLANVGRGVAIRPHADMLILLALYSL
YLFALIVYDSVCMNCGVRSSSYWIFCFASAVHMLVIMSTVAVVPLRLTLTVRISL

```

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best local Similarity:	Query Match:	DB:
561	9.00	100.00%	100.00%	4.23%	3
Length:	260624	Matches:	9	Mismatches:	0
		Conservative:	0	Indels:	0
				Gaps:	0

US-10-024-955-7 (1-213) x AE003625 (1-260624)

Qy 8 AlalayaAlaIaphaValaIaIaIaSer 16
 Db 120532 GCACGAGTCGCATTGTGGCTGTCT 120558

RESULT 24
 AX028708 LOCUS 24 bp DNA linear PAT 16-SEP-2000
 DEFINITION Sequence 12 from patent EP1018550.
 ACCESSION AX028708
 VERSION AX028708.1 GI:10189821
 KEYWORDS
 SOURCE
 ORGANISM
 European house dust mite.
 Dermatophagoides pteronyssinus
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;
 Dermatophagoides.
 1 (bases 1 to 24)
 Thomas, W.R. and Chua, K.Y.
 Allergenic protein and peptides from house dust mite and uses
 thereof
 Patent: EP 1018550-A 12 12-JUL-2000;
 INST CHILD HEALTH RESEARCH (AU)

REFERENCE
 AUTORS
 TITLE
 JOURNAL

FEATURES Location/Qualifiers
source 1..24
/organism="Dermatophagoides pteronyssinus"
/db_xref="taxon:6956"
BASE COUNT 5 a 5 c 4 g 10 t
ORIGIN

Alignment Scores:
Pred. No.: 1.77 Length: 24
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 6 Gaps: 0

US-10-024-955-7 (1-213) x AX028708 (1-24)

QY 173 SerilleleuaspProilephecly 180
|||||
Db 1 TCATCTTGATCCAAATTTTGCC 24

RESULT 25
FCSHZ5R1 205 bp ss-RNA linear VRL 02-AUG-1993
LOCUS FCSHZ5R1
DEFINITION Feline sarcoma virus (H25-Pesv) recombination site DNA.
ACCESSION M14290
VERSION M14290.1 GI:323885
KEYWORDS retrovirus.
SEGMENT 1 of 2
SOURCE Feline sarcoma (H25) Integrated viral DNA.
ORGANISM Feline sarcoma virus
VIRUSES: Retroid viruses; Retroviridae; Mammalian type C
retroviruses; 1-Mammalian type C virus group.
1 (bases 1 to 205)
Besser, P., Lader, E., George, P. C., Bergold, P. J., Qiu, F. H.,
Zuckerman, E. E. and Hardy, W. D.
A new acute transforming feline retrovirus with fms homology
specifies a C-terminally truncated version of the c-fms protein
that is different from SM-feline sarcoma virus v-fms protein
JOURNAL J. Virol. 60 (1), 194-203 (1986)
MEDLINE 86308237
PUBMED 3018286

FEATURES Location/Qualifiers
source 1..205
/organism="Feline sarcoma virus"
/db_xref="taxon:11772"
CDS <1..>205
/note="p30/p10/5'-v-fms coding region"
/codon_start=1
/protein_id="AAA43043.1"
/db_xref="GI:554646"
/translation="HKEMTKYLAIVAAQNRKDRERMPGSGHVGASAEYPPGCPPL
CPASSCCPPTMGFRALLVILVA"
24..25
/note="p30/p10 protein junction"
misc_feature 44 a 63 c 58 g 40 t
BASE COUNT 44 a 63 c 58 g 40 t
ORIGIN

Alignment Scores:
Pred. No.: 12 Length: 205
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 14 Gaps: 0

US-10-024-955-7 (1-213) x FCSHZ5R1 (1-205)

QY 196 LysGLuMetThrIysValIeuAla 203
|||||
Db 4 AAGGAGATGACTMAAGTCTGGCC 27

RESULT 26

AY004870/c
LOCUS AY004870 1009 bp mRNA linear VRT 20-JUL-2000
DEFINITION Gallus gallus coiled-coil domain protein D53 (D53) mRNA, complete
cds.
ACCESSION AY004870
VERSION AY004870.1 GI:9509277
KEYWORDS
SOURCE Gallus gallus.
ORGANISM Gallus gallus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1009)
REFERENCE Proux, V. A., Calochy, G. and Marx, M.
chicken homolog of human D53
TITLE Unpublished
JOURNAL 2 (bases 1 to 1009)
REFERENCE Proux, V. A., Calochy, G. and Marx, M.
Direct Submission
AUTHORS Submitted (15-JUL-2000) Biology, CNRS, Institut Curie, bat 110
JOURNAL Campus Universitaire, Orsay 91405, France

FEATURES Location/Qualifiers
source 1..1009
/organism="Gallus gallus"
/db_xref="taxon:9031"
gene 1..1009
/gene="D53"
1..633
/gene="D53"
/note="similar to Homo sapiens D53"
/codon_start=1
/product="coiled-coil domain protein D53"
/protein_id="AA087084.1"
/db_xref="GI:9509278"
/translation="WERQARGLLETQSLHEDEDWTVDTSMISEEKEELKAE
LAEIDISPLROYLAKEKHLIEIKQKLSMLNELKQFNSKMDHMQTSAYKKTHT
LSHAKQKATAISVNGTAISKFGDRSHSISYSIRHSISMPAMNSPTFYSFEEKVE
TTVTSIKTKVGTSHSGSFEEVLSSTHAAQSSLAGTRLPSESEELQC"

BASE COUNT 292 a 236 c 225 g 256 t
ORIGIN

Alignment Scores:
Pred. No.: 49.9 Length: 1009
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 5 Gaps: 0

US-10-024-955-7 (1-213) x AY004870 (1-1009)

QY 5 leuLeuIIeAlaAlaValAlaIaphe 12
|||||
Db 347 TTGCTGATAGCTGCTGTCTTTT 324

RESULT 27
AY060711/c
LOCUS AY060711 1056 bp mRNA linear INV 08-NOV-2001
DEFINITION Drosophila melanogaster GH15272 full length cDNA.
ACCESSION AY060711
VERSION AY060711.1 GI:16768079
KEYWORDS FLI cDNA.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1056)
REFERENCE Stapleton, M., Broksstein, P., Hong, L., Agdayani, A., Carlson, J.,
Champe, M., Chavez, C., Dorsett, V., Fatfan, D., Frise, E., George, R.,
Gonzalez, M., Guarin, H., Li, P., Liao, G., Miranda, A., Mungall, C. J.,
Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K.,
Yu, C., Lewis, S. E., Rubin, G. M. and Ceiniker, S.

TITLE Direct Submission
JOURNAL Submitted (30-OCT-2001) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA
COMMENT Sequence submitted by: Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720
 This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES
source
 1..1056
 /location/Qualifiers
 /organism="Drosophila melanogaster"
 /strain="y: cn bw sp"
 /db_xref="taxon:7227"
 /map="7A4-7A4"
 1..1056
 /gene="CG1999"
 /note="alignment with genomic scaffold AE003440"
 /db_xref="FLYBASE:FBgn0029947"
 113..745
 /gene="CG1999"
 /note="Longest ORF"
 /codon_start=1
 /product="GH15272P"
 /protein_id="AAL28259.1"
 /db_xref="GI:16768080"
 /db_xref="FLYBASE:FBgn0029947"
 /translation="MQRSSYPICHTVPSLTGLGGMDLRRTMASDSMGREGDNGSPNSPRAGVSASATSTVISAASSTSRKVNAGAEERRLRNEDNMRRDSYIDTKLNRDPNAVRNFRQTEPTSLRKQFMRSPDEISREVMGRDETVRYKRNAAQSKHSVARTERSSDNTNRNQHILQLHMQQQQRCQQQHKKNOQYNNNGRSVESPEDQ"
BASE COUNT 350 a 221 c 232 g 253 t
ORIGIN

Alignment Scores:
 Pred. No.: 52 Length: 1056
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 3 Gaps: 0

US-10-024-955-7 (1-213) x AY060711 (1-1056)
 Oy 8 ALaAlaValAlaPheValAlaVal 15
 Db 677 GCTGCTGTTCCTTGTGTGCTGT 654

RESULT 28
LOCUS AX123559 1128 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 3475 from Patent EP1108790.
ACCESSION AX123559
VERSION AX123559.1 GI:14041047
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae;

REFERENCE 1 (bases 1 to 1128)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K., Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 3475 20-JUN-2001;
 KYOMA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source
 1..1128
 /location/Qualifiers
 /organism="Corynebacterium glutamicum"
 /db_xref="taxon:1718"
BASE COUNT 286 a 292 c 312 g 238 t
ORIGIN

Alignment Scores:
 Pred. No.: 55.2 Length: 1128
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 6 Gaps: 0

US-10-024-955-7 (1-213) x AX123559 (1-1128)
 Oy 8 ALaAlaValAlaPheValAlaVal 15
 Db 38 GCGGCACTTGCTTGTGCTGTC 15

RESULT 29
LOCUS AF272757/C 1220 bp mRNA linear PLN 18-JAN-2002
DEFINITION Zea mays kinesis heavy chain (KIN11) mRNA, partial cds.
ACCESSION AF272757
VERSION AF272757.1 GI:15208462
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 1220)
 Lawrence,C.J., Mainberg,R.L., Muszynski,M.G. and Dave,R.K.
 Maximum likelihood methods reveal conservation of function among closely related Kinesis families
 J. Mol. Evol. 54 (1), 42-53 (2002)
 21592582
 11734897
REFERENCE 2 (bases 1 to 1220)
AUTHORS Lawrence,C.J., Mainberg,R.L., Muszynski,M.G. and Dave,R.K.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2000) Botany, University of Georgia, Miller Plant Sciences Building, Agriculture Drive, Athens, GA 30602, USA
FEATURES
source
 1..1220
 /organism="Zea mays"
 /db_xref="taxon:4577"
 /tissue_type="tassel apical meristem, floral transition"
 <1..1220
 /gene="KIN11"
 <1..1065
 /gene="KIN11"
 /note="C-terminal family member similar to Saccharomyces Kar3p"
 /codon_start=1
 /product="kinesis heavy chain"
 /protein_id="AAK1820.1"
 /db_xref="GI:15208463"
 /translation="VQSAIDGVVCIFFAYGQSGKTYTMGNGPELEDQKMTIPRSLEQFOASOTLNSGGRYKMOASMLEIYNETIRDLATNRMAAOVGAASKYIKDITGNNTNSDLITIDVTISINEYSILRRAAOSRSVGRPMDESSRCFTLRIFGVNEGDOQVQGVNLIDIDLAGSERLNSGATGDRKETAIVAINSSLSLSVYISIAKREHYVFRNSKUTYLLQPCITGDSKTLTFVNLSPASSTGSLSLRFAAVNCSGSSPGAKPRC GAHNDEIRRRRAASVMDICIPISGGVSCQFPIMVALLIALTAGSGCAGLILIL

BASE COUNT 327 a 244 c 312 g 337 t
ORIGIN GPDLICMLGCVAFIFCLHL"

Alignment Scores:
Pred. No.: 59.2 Length: 1220
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x AF272757 (1-1220)

OY 86 ArgGlnGlyAspAlaAsnVallys 93
DB 900 AGACAAGGGAGTCAAAATGTCAAG 877

RESULT 30
E05296 1239 bp DNA linear PAT 29-SEP-1997
LOCUS DNA sequence encoding Methylobacillus glyco genes homoserine
DEFINITION dehydrogenase.
ACCESSION E05296
VERSION E05296.1 GI:2173486
KEYWORDS JP 1993207886-A/2.
SOURCE Methylobacillus glyco genes.
ORGANISM Methylobacillus glyco genes
Bacteria; Proteobacteria; beta subdivision; Methylophilus group;
Methylobacillus.
REFERENCE 1 (bases 1 to 1239)
AUTHORS Motoyama, H., Anazawa, H., Ishino, S. and Teshiba, S.
TITLE PRODUCTION OF L-THREONINE BY FERMENTATION METHOD
JOURNAL Patent: JP 1993207886-A 2 20-AUG-1993;
KIOWA HAKKO KOGYO CO LTD
OS Methylobacillus glyco genes
PN JP 1993207886-A/2
PD 20-AUG-1993
PF 29-JAN-1992 JP 1992014335
PI MOTONYAMA HIROAKI, ANAZAWA HIDEKI, ISHINO SHUICHI, PI TESHIBA
SADAO
PC C12P13/08, (C12P13/08, C12R1:07);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain-1006;
CC *source: deposit-ATCC21371;
FH Key Location/Qualifiers
FT CDS 1..1239
FT /product='homoserine dehydrogenase',
FT /organism='Methylobacillus glyco genes',
FT /db_xref='taxon:406'
BASE COUNT 272 a 318 c 351 g 298 t
ORIGIN

Alignment Scores:
Pred. No.: 60 Length: 1239
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x E05296 (1-1239)

OY 34 AspAspAlaIleAlaIleGlu 41
DB 1156 GATGATGCCATCGCCGATTCGAG 1179

RESULT 31
CGRECAG/C 1324 bp DNA circular BCT 28-JUN-1995
LOCUS C.glutamicum recA gene.
DEFINITION X77384
ACCESSION X77384.1 GI:475063
VERSION recA: recA gene.
KEYWORDS Corynebacterium glutamicum.
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1 (bases 1 to 1324)
AUTHORS Billman-Jacobe, H.
TITLE Nucleotide sequence of a recA gene from Corynebacterium glutamicum
JOURNAL DNA Seq. 4 (6), 403-404 (1994)
MEDLINE 95143581
PUBMED 7841463
REFERENCE 2 (bases 1 to 1324)
AUTHORS Billman-Jacobe, H.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1994) H. Billman-Jacobe, CSIRO Division of Animal
Health, Private Bag No. 1, Cnr. Park Drive and Flemington Rd,
Parkville 3052, AUSTRALIA
revised by [3] MAT
REFERENCE 3 (bases 1 to 1324)
AUTHORS Billman-Jacobe, H.
TITLE Direct Submission
JOURNAL Submitted (26-APR-1994) H. Billman-Jacobe, CSIRO Division of Animal
Health, Private Bag No. 1, Cnr. Park Drive and Flemington Rd,
Parkville 3052, AUSTRALIA
On Apr 27, 1994 this sequence version replaced gi:453192.
COMMENT Location/Qualifiers
FEATURES
SOURCE
1..1324
/organism='Corynebacterium glutamicum'
/strain='AS019'
/db_xref='taxon:1718'
136..1266
/gene='recA'
136..1266
/gene='recA'
/codon_start=1
/transl_table=1
/protein_id='CA54563.1'
/db_xref='GI:475064'
/db_xref='SWISS-PROT:P42442'
/translation='MHPKTKATKATKAGNDROKALDAIALLEKDFGKCAVRLGDE
NRPIQFISSGNTAIDIALIGFPPRGRIYEVGESSGTTVALHAIAQAQAGGIA
AFIDAEHALDPDYARKLGVDLDALVQPTGEQALFIADMLVSGAIDIIIVDSVA
LTPRAIEGEGDGHVGLQARLMSQALRKMGALYNSGTAIFINQDREKIGVFGSP
ETTTGKALKFVYASVRCDDIRIOTLIXGQDAIGNRRLKYVKNKVSPPRIAEFDIMY
GEGISRESSVIDLAVDNGIVYKSGSMFTYGEQDGGKREYVRLSLKINPELTDELEK
IFKRLGKTAASDELTDPELVLPVNDVDDADTDAED'
BASE COUNT 352 a 331 c 351 g 290 t
ORIGIN

Alignment Scores:
Pred. No.: 63.7 Length: 1324
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x CGRECAG (1-1324)

OY 8 AlaAlaValAlaIleValAlaVal 15
DB 173 GCGGACGTTGCCCTTGTCGTGTC 150

RESULT 32
CGU14965/C 1402 bp DNA linear BCT 05-FEB-1999
LOCUS Corynebacterium glutamicum RecA (recA) gene, complete cds.
DEFINITION

ACCESSION U14965
 VERSION U14965.1 GI:556615
 KEYWORDS
 SOURCE Corynebacterium glutamicum.
 ORGANISM Corynebacterium glutamicum
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacteriaceae; Corynebacterium.
 REFERENCE 1 (bases 1 to 1402)
 AUTHORS Kerins, S.M., Fitzpatrick, R., O'Donohue, M., and Dunican, L.
 TITLE Molecular cloning and characterization of the recA gene from
 Corynebacterium glutamicum AS019
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1402)
 AUTHORS Kerins, S.M.
 TITLE Direct Submission
 JOURNAL Submitted (21-SEP-1994) Sinead M. Kerins, Department of
 Microbiology, University College, Galway, Ireland
 FEATURES
 source 1..1402
 /organism="Corynebacterium glutamicum"
 /strain="AS019"
 /db_xref="taxon:1718"
 /clone="PSK1"
 133..1263
 /gene="recA"
 133..1263
 /gene="recA"
 /codon_start=1
 /transl_table=11
 /product="RecA"
 /protein_id="AAD12743.1"
 /db_xref="GI:556616"
 /translation="MAPKKTATKATKAANDROKALDALALAIENKDFGKAVMRLEDE
 NRPPIOTISSGTAIDIALIGGFPGRGRVEYGGESSGKTVALHAIQAQKGGIA
 AFIDAEHADDPYARKIGVDPDALLYSPDGEQALETADMVRSAGADIIYIDSVAA
 LTPKAEIEMSDSHVGLQARLMSQALRKMTALVNSGTTAIFINOLRKIGVMFGSP
 ETTTGKALKETASVCDIRIQTLDGDGDAIGNRRLKVVNKKVSPPKIAEFIMY
 GGGISRESSVIDLAVNGIIVKSGSMFTYEGGLOGKKEKRLSKENPELDELEDK
 IFKKLGVGKYAASDELTFDDPEVLVPNVDFDEADTEADAE"
 BASE COUNT 380 a 350 c 369 g 303 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 67 Length: 1402
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 1 Gaps: 0
 US-10-024-955-7 (1-213) x CGU14965 (1-1402)
 QY 8 AAlaIaVaJaIaPhVeJaIaIaVal 15
 Db 170 GCGGAGTTCCTTGTGTCTGTC 147
 RESULT 33 AF504294 1590 bp mRNA linear VRL 19-MAY-2002
 LOCUS Nairobi sheep disease virus isolate Ig619 nucleocapsid mRNA,
 DEFINITION complete cds.
 ACCESSION AF504294
 VERSION AF504294
 KEYWORDS AF504294.1 GI:20977202
 SOURCE Nairobi sheep disease virus.
 ORGANISM Nairobi sheep disease virus.
 REFERENCE 1 (bases 1 to 1590)
 AUTHORS Viruses; ssRNA negative-strand viruses; Bunyaviridae; Nairovirus.
 TITLE Marzinko, B.I. and Nichol, S.T.
 JOURNAL Nairobi sheep disease virus, an important tick-borne pathogen of
 sheep and goats in Africa, is also present in Asia
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1590)
 AUTHORS Marzinko, B.I. and Nichol, S.T.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-2002) Special Pathogens Branch, CDC, 1600 Clifton
 Rd. NE., Atlanta, GA 30333, USA
 FEATURES
 source 1..1590
 /organism="Nairobi sheep disease virus"
 /isolate="Ig619"
 /db_xref="taxon:194540"
 /country="India"
 /note="Asian variant of Nairobi sheep disease virus
 antigenomic sense S segment sequence
 synonym: Ganjam virus; segment: S"
 52..1500
 /note="encapsulation of viral RNAs"
 /codon_start=1
 /product="nucleocapsid"
 /protein_id="AAM33324.1"
 /db_xref="GI:20977203"
 /translation="MNOIVADNKDAIILAMHKTYSKHKLSVLTSASFCEITPDLS
 GVEYSMLRVSSSEKDSYASALVATKFCAPILIECAWTSCTGMIORELMDNNGEM
 VKIMVADYGLKLTETVSPPEOLLGYORALKRKDKTKVGINMTALAAKIAFDEVKPG
 SIIVVVKDMLSMITRRNKILNROSSEDEVPRKPKSKHIDMARDLAQKFLYVNP
 WEDINKAGKSGIALATGMAKILIEDGKRVADDEKESLSLVAMINAKDEYENGKEV
 VDGIRKHLQKALELAKQSSAMRAQAQIDTVFSSYVWLKAGVTAEMPTVSQFLFEL
 GKVPGRNKRKMKALSMPILKWKGLLAFADDSFTANRIYMHGVLITAGMSSEIGVCF
 GAIPVANDDAEGSGHINKILNTYTDQAGNPCAQNIVALFNIOKAGFIJESMDIYA
 SEHLHQSIVGCKRSPFNAYNIRGNATSIQII"
 BASE COUNT 451 a 356 c 404 g 379 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 75 Length: 1590
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 14 Gaps: 0
 US-10-024-955-7 (1-213) x AF504294 (1-1590)
 QY 176 AspproTlePhcGlyValIeuser 183
 Db 474 GATCCCATATTGGTGTCTTCT 451
 RESULT 34 AF265671 1827 bp mRNA linear VRT 07-AUG-2000
 LOCUS Xenopus laevis paired-like homeodomain transcription factor Ptx3
 DEFINITION mRNA, complete cds.
 ACCESSION AF265671
 VERSION AF265671.1 GI:9716649
 KEYWORDS
 SOURCE Xenopus laevis.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 1827)
 AUTHORS Khosrowshahian, F., Chang, W.Y., Lum, N., Fujiki, K. and Crawford, M.J.
 TITLE xPtx3 exhibits dynamic expression during sensory placode
 development in Xenopus
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1827)
 AUTHORS Khosrowshahian, F., Chang, W.Y., Lum, N., Fujiki, K. and Crawford, M.J.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-2000) Biological Sciences, University of Windsor,
 401 Sunset Location/Qualifiers
 FEATURES
 source 1..1827
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"

CDS

158..1036
/codon_start=1
/product="paired-like homedomain transcription factor
p1ex3"
/protein_id="AA097592.1"
/db_xref="GI:9716650"
/translation="MDPNLLTDESEARSPALSLSDSGTPQHDHSCGQEHSDTEKSOON
QTDSDNEDGILKKRORRTHFTSOOLELEAFOHNRYPDMSTREELVMTLTA
RVRVFKNRRAKMRKRRNOAEICNSFGQFGLMOPDDMSSGYNMATAKGLA
TSPLSAKSPQEFNSMANTSPSSQPEPSPSISTMTNTSSKVPASVTVGPGSSINNG
NINNLNPSLTSASASACPASTASPYMRDTCNSSLASLRLKAKOHANFTYPAVOT
PASNLSPCQYAVDRPV"

BASE COUNT 593 a 414 c 339 g 480 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 84.9 Length: 1827
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 5 Gaps: 0

US-10-024-955-7 (1-213) x AE265671 (1-1827)

Oy 205 Alaphelysargylulegulus 212
|||||

Db 1781 GCCTTTAAAGGAATTGGAAAA 1758

RESULT 35
AX244529 1833 bp DNA linear PAT 28-SEP-2001
LOCUS Sequence 31 from Patent W00166568.
DEFINITION AX244529
ACCESSION AX244529
VERSION AX244529.1 GI:15859456
KEYWORDS
SOURCE
ORGANISM
Feline leukemia virus.
Viruses: Retrovirus: Retroviridae: Gammaretrovirus.
REFERENCE
1 (bases 1 to 1833)
Jensen, W.A., Lappin, M.R., Rosen, D.K. and Andrews, J.S.
Use of recombinant antigens to determine the immune status of an
animal
Patent: WO 0166568-A 31 13-SEP-2001;
Heska Corporation (US); COLORADO STATE UNIVERSITY RESEARCH
FOUNDATION (US)
JOURNAL
Location/Qualifiers
FEATURES
source
1..1833
/organism="Feline leukemia virus"
/db_xref="taxon:11768"
1..>1833
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC88571.1"
/db_xref="GI:15859457"
/translation="MEHLRLYLRLAGLGAARHPNLLAOVKQFLOGKEETPASFLE
RIKEAYMTPYDPEDEGQASVLSFIYSSDIPRKRLRLBGLQFTSLDLKEKE
KITNKRTPPERERELMROEBRDKHKHETKIVLAVONRNDREESKLGQRKI
PLGDOCAVCEKRGHWDCPNRPRKRPANSTLLNLDMNPSPOMYNVTWITNO
TNTQANATSMGLTLDVPTLHYDLCDLVDTEPMVLSPTGYPPSYGCKTIDRRKO
QOYVPEYVCGHPSILGKTHCGAGDGCACAGCETGCEAMMKPSSMDYITVRKG
SSONNNEGKCNPLILOFTQKQASWDGPMKMLRLRYFTGYDPTIALFTSRRTT
PPAMGPDVLPOKPPRSQSGSKATOPNESAPRSVAPPTGPKRKGDLL
INLVQAYLALNMTDPRKTKDCMLCLVSRPYEGILILNYSQMTNPSPSCSIPPH
KLITISKVGGGLGTVKTHQALCNKTHGADYRAAPRYLAANGTYMACNTGL
TPCISMAVNLITSDPCVLIELMPRYTHQBEYVYTHRAKGRFR"

BASE COUNT 551 a 479 c 430 g 373 t

ORIGIN

Alignment Scores:
Pred. No.: 85.2 Length: 1833
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 6 Gaps: 0

US-10-024-955-7 (1-213) x AX244529 (1-1833)

Oy 196 Lysglumetthylusvalleuola 203
|||||

Db 391 AAGGAGATGACTAAAGTTCTGCCC 414

RESULT 36
STMHRDC 1965 bp DNA linear BCT 26-APR-1993
LOCUS Streptomyces aureofaciens principal sigma- like factor (hrdc) gene,
DEFINITION complete cds.
ACCESSION M90412
VERSION M90412.1 GI:153308
KEYWORDS principal sigma-like factor; rpoD homologue.
SOURCE Streptomyces aureofaciens (strain CCM 3239) DNA.
ORGANISM Streptomyces aureofaciens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomycetes.
REFERENCE
1 (bases 1 to 1965)
Kormanec, J., Farkasovsky, M. and Potuckova, L.
Four genes in Streptomyces aureofaciens containing a domain
characteristic of principal sigma factors
JOURNAL Gene 122 (1), 63-70 (1992)
MEDLINE 93083996
PUBMED 1452038

FEATURES
source
location/Qualifiers
1..1965
/organism="Streptomyces aureofaciens"
/strain="CCM 3239"
/db_xref="taxon:1894"
37..1779
/gene="hrdc"
37..69
/gene="hrdc"
/rpt_type="inverted
114..119
/gene="hrdc"
126..1712
/gene="hrdc"
/gene="hrdc"
/codon_start=1
/transl_table=1
/product="principal sigma-like factor"
/protein_id="AA26764.1"
/db_xref="GI:153309"
/translation="MSLSPTSPPTPEIAESEALVALVERGREGQHTINGDDVROAFEG
RIPVDOKRVLRLSLNOYLDEEGVALIHVSAPATKRAAKKRRKAAPAPRTVTKAAP
PRYIGARKTSAAPRAKAAIISAPSAATGDEATREAAAEPRKRTVTKTAKKATATK
TAATKRGAGATGDAENPAVEGEDMAVEDLADAEETPRAGTQGVLSDDDDDAAP
AQGVVAGATADAPVKDYLKLGIVPLNABQVELEARIAGLSEKYLEEDHRA
FKRELELIVDEGRRKNNHLEANLRLVNSLAKRYTGCMFLDLIOGNGLIRAVK
EDYTKGKFTSTVATWMLROAITRAMADOOSTIRIPYHVEIINKLARVOROMLDIGR
EPPEELGKFLDMPREKVIIVOKYGRPISLHNPIDEGGSEEGDILDESEAVVPADA
VSTFTLEDOQLSLGLTISEEAGVSKRYGLNQGPTLDEIGRYGVTRERIRQIES
KTKMSKLNHPSRSQLRGYLD"

repeat_unit
/gene="hrdc"
/rpt_type="inverted
457..483
repeat_region
/rpt_type="inverted
1753..1779
repeat_region
/rpt_type="inverted
BASE COUNT 337 a 721 c 658 g 249 t

ORIGIN

Alignment Scores:
Pred. No.: 90.6 Length: 1965
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.768 Indels: 0
 DB: 1 Gaps: 0

US-10-024-955-7 (1-213) x STMHRDC (1-1965)

OY 204 ProAlaphelysarglueglu 211
 DB 948 CCCGCTCAAGCGGAGCTGAG 971

RESULT 37
 LOCUS 001047/c 2054 bp mRNA linear VRT 15-JUN-1994
 DEFINITION Gallus domesticus RAD52 homolog mRNA.
 ACCESSION 001047
 VERSION 001047.1 GI:437100
 KEYWORDS
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 2054)
 Bezuubova,O.Y., Schmidt,H., Ostermann,K., Heyer,W.D. and
 Buerstedde,J.W.
 Identification of a chicken RAD52 homologue suggests conservation
 of the RAD52 recombination pathway throughout the evolution of
 higher eukaryotes

REFERENCE
 AUTHORS Nucleic Acids Res. 21 (25), 5945-5949 (1993)
 MEDLINE 94119695
 PUBMED 8290357
 2 (bases 1 to 2054)
 Bezuubova,O.O.
 Direct Submission
 TITLE Submitted (25-AUG-1993) Olga O.Y. Bezuubova, Basel Institute for
 JOURNAL Immunology, Grenzacherstr. 487, Basel, Basel city, CH-4005,
 Switzerland
 FEATURES
 Location/Qualifiers
 source 1..2054
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /tissue_type="testis"
 BASE COUNT 606 a 496 c 519 g 433 t
 ORIGIN

Alignment Scores:
 Pired. No.: 94.3 Length: 2054
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 5 Gaps: 0

US-10-024-955-7 (1-213) x 001047 (1-2054)

OY 11 AlAphEVALAlAYAlSerAlasp 18
 DB 836 GCTTTGTGCGAGTTCCGCTGAT 813

RESULT 38
 LOCUS FCVCP 2565 bp ss-RNA linear VRL 13-APR-2000
 DEFINITION feline leukemia virus long terminal repeat, complete sequence; gag
 (gag) gene, complete cds, alternatively spliced; and gag/pol
 polyprotein (gag:pol) gene, partial cds.
 K01803
 K01803.1 GI:323914
 VERSION
 KEYWORDS DNA polymerase; gag protein; pol polyprotein; polymerase.
 SOURCE feline leukemia virus.
 ORGANISM feline leukemia virus
 Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
 1 (bases 1 to 2565)
 Lapievotte,I., Hampe,A., Sherr,C.J. and Galibert,F.
 Nucleotide sequence of the gag gene and gag-pol junction of feline

JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Leukemia virus
 J. Virol. 50 (3), 884-894 (1984)
 84216470
 6328019
 [1] compared the Felv with two strains of the feline sarcoma virus
 and with that of the Moloney strain of murine leukemia virus. A
 high degree of homology was found.
 The predicted structure of the complete Felv gag gene precursor
 suggests that the translation of non-glycosylated (gag-Pr65) and
 glycosylated (gag-gp80) gag gene polyproteins is initiated at two
 different but in-phase start codons, which are separated by a 222
 bp segment which can encode an amino terminal 7.6 kd peptide.

FEATURES
 source
 LTR
 gene
 cps

1..2565
 /organism="Feline leukemia virus"
 /db_xref="taxon:11768"
 /note="serogroup: B"
 1..244
 /standard_name="5' LTR"
 445..2187
 /gene="gag"
 445..2187
 /gene="gag"
 /gene="gag"
 /standard_name="gag-gp80"
 /codon_start=1
 /protein_id="AAA43054.1"
 /db_xref="GI:323915"
 /translation="MSGASGTAIGAHIFGVSPERYVLISGSGAGPSKSLSEVPSVW
 YRSRAKVIICLVASPIVPCITPLIAEAVNGQVYTPPLSLTDHMSVRAHNGV
 EVRKRWITLCEAEWMMVGVMPREGTSLDNISOVERKIFAPGPHGHDVPIYITW
 RSLATDPPSWAPFLPPKPPPTPLPQPLSPASAPLSSLPVVPKDPPLPVLPP
 PSSPLIDLTEEPPEPGHGPPSGPTPASPASPLASRLRRENPAPESQALLREG
 PNNRPQWPFASDLYNMKSHNPSPQVALTNIESILVTHQPTWDCQALLQALL
 TGEERQVILLEARKOVPEGDRPQLPVVIDETPLTPRNMDFPATPAGREHLRLYROL
 LLAGLRGAARPTNLAQYKOYVOCGEETPAFLERIKAYMYRMPYDEPDGCAASYI
 LSTYQSSPDIRNKLQRLLEGLOGFTLSLDAEAEITIKRTTPPERERRLQROBERD
 KRRHEKMTVLATVVAQNDRKIDRENKLGIDRKIPLGKDCAYCKEKGHWDCPKRP
 RKKPANSSTLNLED"
 667..2187
 /gene="gag"
 /standard_name="gag-Pr65"
 /codon_start=1
 /protein_id="AAA43055.1"
 /db_xref="GI:323916"
 /translation="MGQVTTPTPLSLTDHMSVRAHNGQVYTPPLSLTDHMSVRAHNGV
 MNNVMPREGTSLDNISOVERKIFAPGPHGHDVPIYITWRSIATDPPEVPELP
 PKPPTPLPQPLSPASPLASRLRRENPAPESQALLREGPNNRPQWPFASDLY
 NMSKSHNPSPQVALTNIESILVTHQPTWDCQALLQALLTGEERQVILLEARKOV
 PGEDRPTQLPNVIDETPLTPRNMDFPATPAGREHLRLYROLLLAGLRGAARPTNLA
 QYKOYVOCGEETPAFLERIKAYMYRMPYDEPDGCAASYI LSTYQSSPDIRNKLQ
 RLEGLOGFTLSLDAEAEITIKRTTPPERERRLQROBERD KRRHEKMTVLATVVA
 QNDRKIDRENKLGIDRKIPLGKDCAYCKEKGHWDCPKRP RKKPANSSTLNLED"
 670..1047
 /gene="gag"
 /product="gag-p15 polyprotein"
 1048..1257
 /gene="gag"
 /product="gag-p12 polyprotein"
 1258..2001
 /gene="gag"
 /product="gag-p30 polyprotein"
 2002..2184
 /gene="gag"
 /product="gag-p10 polyprotein"
 2180..2565
 /gene="pol"
 <2180..>2565
 /gene="gag:pol"
 /codon_start=1
 /product="gag/pol polyprotein"
 /protein_id="AAA43056.2"
 /db_xref="GI:7548183"

BASE COUNT 863 a 761 c 586 g 547 t
ORIGIN TWAOSIROACISAA"

Alignment Scores:
Pred. No.: 123 Length: 2757
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 5 Gaps: 0

US-10-024-955-7 (1-213) x AF378334 (1-2757)

Qy 8 AAlaAlaValaAlaPheValaAlaVal 15
Db 916 GCTGCTGTTCATTGTGTCAGTT 893

RESULT 41
LOCUS AR193470 3022 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 8 from patent US 6348196.
ACCESSION AR193470
VERSION AR193470.1 GI:20240062
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3022)
AUTHORS Audonnet,J.-C., Bouchardon,A., Baudu,P. and Riviere,M.
TITLE Feline polynucleotide vaccine formula
JOURNAL Patent: US 6348196-A 8 19-FEB-2002;
FEATURES
source 1.3022
Location/Qualifiers
/organism="unknown"

BASE COUNT 824 a 831 c 714 g 653 t
ORIGIN

Alignment Scores:
Pred. No.: 133 Length: 3022
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 6 Gaps: 0

US-10-024-955-7 (1-213) x AR193470 (1-3022)

Qy 196 LysGluMetThrLysValLeuAla 203
Db 1537 AAGGAGATGACTAAAGTTCTGGCC 1560

RESULT 42
LOCUS A69020 3023 bp DNA linear PAT 06-MAY-1999
DEFINITION Sequence 8 from Patent WO9803660.
ACCESSION A69020
VERSION A69020.1 GI:4759947
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3023)
AUTHORS Baudu,P., Riviere,M., Audonnet,J. and Bouchardon,A.
TITLE FELINE POLYNUCLEOTIDE VACCINE FORMULA
JOURNAL Patent: WO 9803660-A 8 23-JAN-1998;
BAUDU PHILIPPE (FR)
Other Publication FR 2751223 19980123.
COMMENT 1.3023
Location/Qualifiers
/organism="unidentified"
/strain="GLASGOW-1 TYPE A"
/db_xref="taxon:32644"

gene 1.1743
/gene="GAG"
1.1743
CDS /gene="GAG"
/standard_name="GAG"
/codon_start=1
/product="CAPSID PROTEIN PRECURSOR"
/protein_id="CAB42471.1"
/db_xref="GI:4759948"
/translation="MSGASSGGAIGAHLEGVSPERYVLIGDEGAGSRSLSEVPSVW
YORRARLVICIVASFLVPCITFLIAETVSGORTITPTSLIDHWSVRARAHNOGV
EYRKKWITLCEAEVMMNMGHPRGTSISDSISVEKKITAPGYPGHDPYITTH
RSLATDPSPSWARPELPPKPPYLPQLSPQSPALSTSLYPLPKPPKPPYLPD
PSSPLIDLTETPEPPYPGSGPPSPGPTPIASPIASLRERRRNPAAEESQALPREG
PNMRPOYMPFASADLYNMKSHNPFSSQDPVALTNLIESILVTHQPTWDDCOLQALL
TGEERQVLEARKOVPGEDGPTQLPNVITETPLTRPNMDFATPAGREHLRLYROL
LLAGRGARPRPTNLAQKOVYQKEETPAFLEELKEAYMNPYPYDEDDGOASVIL
LSFIYQSSPDITNKIQRLQLEGLOGFTLSDLTLEAKIYKRRRERERERLMOOREED
KRRHKEMTKVLATVVAQNRKDRRESKLGDRKIFLGRKQCAVCCKEKGHWRDCKRP
RKKPANSSTLNLGD"

gene 1736.3022
/gene="POL"
1736.3022
CDS /gene="POL"
/partial
/gene="POL"
/standard_name="POL"
/codon_start=1
/product="PROTEASE"
/protein_id="CAB42472.1"
/db_xref="GI:4759949"

BASE COUNT 824 a 831 c 714 g 654 t
ORIGIN YSLKDGQRMFLTKARKEALISIPVNPMPROVREFLGTA"

Alignment Scores:
Pred. No.: 133 Length: 3023
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 6 Gaps: 0

US-10-024-955-7 (1-213) x A69020 (1-3023)

Qy 196 LysGluMetThrLysValLeuAla 203
Db 1537 AAGGAGATGACTAAAGTTCTGGCC 1560

RESULT 43
LOCUS BD006097 3023 bp DNA linear PAT 31-JAN-2002
DEFINITION Feline polynucleotide vaccine formula.
ACCESSION BD006097
VERSION BD006097.1 GI:18634468
KEYWORDS JP 2001500112-A/8.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3023)
AUTHORS Audonnet,J.-C., Bouchardon,A., Baudu,P. and Rivie,M.
TITLE Feline polynucleotide vaccine formula
JOURNAL Patent: JP 2001500112-A 8 09-JAN-2001;
MERTAL
COMMENT OS Fely-A
PN JP 2001500112-A/8
PD 09-JAN-2001

PF 15-JUL-1997 JP 1998506629
 PI 19-JUL-1996 FR 96/09337
 PI JEAN CHRISTOPHE AUDONNET, ANNABELLE BOUCHARDON, PHILIPPE BAUDU,
 PI MICHEL RIVIERE
 PC C12N15/48, C12N15/35, C12N15/50, C12N15/38, C12N15/40, C12N15/49,
 PC C12N15/47,
 PC A61K39/295
 CC
 FH Key Location/Qualifiers
 FT CDS 1740.
 FEATURES
 source 1..3023
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 824 a 831 c 714 g 654 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 133 Length: 3023
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: Gaps: 0
 US-10-024-955-7 (1-213) x BD006097 (1-3023)
 QY 196 lysglumetrlhlysalenala 203
 Db 1537 AAGGAGATGACTTAAGTCTGCGCC 1560
 RESULT 44
 AK092132 3409 bp mRNA linear PRI 15-JUL-2002
 LOCUS Homo sapiens cDNA FLJ34813 fls, clone NT2NE2007725, highly similar
 DEFINITION to Homo sapiens cell cycle progression 2 protein (CPR2) mRNA.
 AK092132
 VERSION D14071.1 GI:21750650
 ACCESSION
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens testis carcinoma cell_line:NT2 cDNA to mRNA,
 clone_1lb:NT2NE2 clone:NT2NE2007725.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS 1
 Nlinomiyu, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
 Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,
 Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
 Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
 Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
 Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
 Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
 Nagahara, I., Masuho, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 TITLE 2 (bases 1 to 3409)
 JOURNAL Isogai, T. and Yamamoto, J.
 REFERENCE Direct Submission
 AUTHORS Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
 TITLE Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
 JOURNAL (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: HRI and
 RAB; annotation: HRI and RAB.
 FEATURES
 source 1..3409
 Location/Qualifiers
 1..3409
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="NT2NE2007725"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_1lb="NT2NE2"
 /note="cloning vector: PME18SFL3-mRNA from NT2 neuron
 after the differentiation of NT2 neuronal precursor
 cells."
 BASE COUNT 731 a 928 c 948 g 802 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 148 Length: 3409
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: Gaps: 0
 US-10-024-955-7 (1-213) x AK092132 (1-3409)
 QY 5 Leuleu1lealalaalavala1aphe 12
 Db 535 CTGCTTATTGACAGTTCGATTC 558
 RESULT 45
 MT1HTT
 LOCUS MT1HTT 3410 bp DNA linear BCT 20-FEB-1999
 DEFINITION M. glycoenes genes for homoserine dehydrogenase, threonine
 synthase and thymidylate synthase.
 ACCESSION D14071
 VERSION D14071.1 GI:454871
 KEYWORDS homoserine dehydrogenase; threonine synthase; thymidylate synthase.
 SOURCE Methylobacillus glycoenes (strain ATCC21371) DNA, clones pTK-1 and
 pTKS-1.
 ORGANISM Methylobacillus glycoenes
 Bacteria; Proteobacteria; beta subdivision; Methylophilus group;
 Methylobacillus.
 REFERENCE
 AUTHORS 1 (bases 1 to 3410)
 TITLE Motoyama, H., Maki, K., Anazawa, H., Ishino, S. and Teshiba, S.
 Cloning and nucleotide sequences of the homoserine dehydrogenase
 genes (hom) and the threonine synthase genes (thrc) of the
 gram-negative obligate methylotroph Methylobacillus glycoenes
 194, Japan (tel: 0427-25-2555(ex.322), Fax: 0427-26-8330)
 JOURNAL Appl. Environ. Microbiol. 60 (1), 111-119 (1994)
 MEDLINE 94161493
 JOURNAL 2 (bases 1 to 3410)
 REFERENCE Motoyama, H.
 AUTHORS Direct Submission
 TITLE Submitted (18-JAN-1993) Hiroaki Motoyama, Kyowa Hakko Kogyo Co.,
 JOURNAL Ltd., Tokyo Research Laboratories; 3-6-6 Asahimachi, Machida, Tokyo
 194, Japan (tel: 0427-25-2555(ex.322), Fax: 0427-26-8330)
 FEATURES
 source 1..3410
 Location/Qualifiers
 1..3410
 /organism="Methylobacillus glycoenes"
 /strain="ATCC21371"
 /db_xref="taxon:406"
 /clone_1lb="DH5alpha"
 399..405
 491..1729
 491..1729
 /gene="hom"
 /gene="hom"
 /EC_number="1.1.1.3"
 /codon_start=1
 /transl_table=11
 /product="homoserine dehydrogenase"
 /protein_id="BAA40415.1"
 /db_xref="GI:2160268"
 /translation="MKPINVGLIGTVGGSTYVLPTRNOEIRARBRGPAITRVAD
 RNLEIARVYTGKIDYDDDAIYSDPAIDIVELIGSYTVAARELYLKAIEGKHYVT
 ANKALICAMAKFLPLRRKKAASSLLKPLLVSVLFRAVREGLAANIEVITAGING
 TTNFTLSEMRREGIAFADVLKAEORLGYAEADPFEDVGGIDAARKLMTLMTLVFVH
 SLRGITKIDAVDITKRPDKGVLELRVHPTLIPERKLICQCEMRNECCAGGRCWPTL

YGAGAGAEPTASAVADLVGTDRGISCPHLAFGQDRVLVDLPILPIGETISSAYVRLR
 AVDKPGVLADVTRILGRQISIDAMIQKPEQGEQDAIILLTHVTVEKNMDAIAAI
 EALPAISGKVTIRMEELSR"
 1779. 3206
 /gene="thrc"
 1779. 3206
 /gene="thrc"
 /EC_number="4.2.99.2"
 /codon_start=1
 /transl_table=11
 /product="threonine synthase"
 /protein_id="BAA0416.1"
 /db_xref="GI:2160269"
 /translation="MKYISTRGQSPALSFSEIILGLADPGGLVPEQYQPSADALS
 ARGMNVRDLAFTILSLRILDDIPADDLRIIVKTYRADVYAYARPGQDAEDITPYKL
 EDDLYLISNGPTLAFKDMAMQLNGEYVLAQGGETTIIIGATSGDTSAAEYAM
 RGKQGVKFMLSPHOKMSRQTAQMSLDDNIPINAVKVGPDQDQIVKAVSNHAF
 KAKNNTIGAVNSINMARVAQVYYTYKRGTYAVADNAQVSPVSPSNGNVCAGHAR
 MNGLPKLVATNENDVLDFFKTVTRPGSANTYHTSSPMDSKSNFEREVED
 LVGRDAKRYRELKMGVDAGSFDLNDGMPFAYADYGFVSGSSNHANMQTKATHER
 YGVITDHTADGLKVALEHREAGTPMLVETALPAKFDAIVEALGHKPERPSLEGL
 ESLPQFEVMEADAIVIKQFIVEHI"
 3219. 3410
 /gene="thya"
 3219. 3410
 /gene="thya"
 /EC_number="2.1.1.45"
 /codon_start=1
 /transl_table=11
 /product="thymidylate synthase"
 /protein_id="BAA03157.1"
 /db_xref="GI:2160270"
 /translation="MKVYHDLMRHVLHGHKKEDRTGTGLSVFGYQMRFDLAECPPL
 LTTKKVHLSTIIHLLMFLQ"
 773 t
 BASE COUNT 755 a 924 c 958 g 773 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 148 Length: 3410
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 1 Gaps: 0
 US-10-024-955-7 (1-213) x MTLHTT (1-3410)
 QY 34 ASPASPAIAIleAlaIleAlaIleGlu 41
 ||||||||||||||||||||||||||||
 Db 1646 GATGATGCCATCGCCGCGCATTTGAG 1669
 RESULT 46
 ARO11451
 LOCUS ARO11451 3674 bp DNA linear PAT 04-DEC-1998
 DEFINITION Sequence 324 from patent US 5762938.
 ARO11451
 ACCESSION ARO11451
 VERSION ARO11451.1 GI:3969441
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 REFERENCE 1 (bases 1 to 3674)
 AUTHORS Paolletti,E., Perkins,M.E., Taylor,J., Tartaglia,J., Norton,E.K.,
 Riviere,M., de Taisne,C., Limbach,K.J., Johnson,G.P., Pincus,S.E.,
 Cox,W.I., Audonnet,J.-C.,Francis, and Gettig,R.Robert.
 Modified recombinant vaccinia virus and expression vectors thereof
 Patent: US 5762938-A 324 09-JUN-1998;
 TITLE Location/Qualifiers
 JOURNAL 1. 3674
 FEATURES /organism="unknown"
 source 970 a 1018 c 884 g 802 t
 BASE COUNT 970 a 1018 c 884 g 802 t
 ORIGIN

Alignment Scores:
 Pred. No.: 159 Length: 3674
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 6 Gaps: 0
 US-10-024-955-7 (1-213) x ARO11451 (1-3674)
 QY 196 LysGluMethrlyrsVallenaIa 203
 ||||||||||||||||||||||||
 Db 2188 AAGGAGATGACTTAAGTTCTGGCC 2211
 RESULT 47
 I118089
 LOCUS I118089 3674 bp DNA linear PAT 07-OCT-1996
 DEFINITION Sequence 324 from patent US 5494807.
 I118089
 ACCESSION I118089
 VERSION I118089.1 GI:1598444
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 REFERENCE 1 (bases 1 to 3674)
 AUTHORS Paolletti,E., Perkins,M.E., Taylor,J., Tartaglia,J., Norton,E.K.,
 Riviere,M., de Taisne,C., Limbach,K.J., Johnson,G.P., Pincus,S.E.,
 Cox,W.I., Audonnet,J.-C.,Francis, and Gettig,R.R.
 NYVAC vaccinia virus recombinants comprising heterologous inserts
 Patent: US 5494807-A 324 27-FEB-1996;
 TITLE Location/Qualifiers
 JOURNAL 1. 3674
 FEATURES /organism="unknown"
 source 970 a 1018 c 884 g 802 t
 BASE COUNT 970 a 1018 c 884 g 802 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 159 Length: 3674
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 6 Gaps: 0
 US-10-024-955-7 (1-213) x I118089 (1-3674)
 QY 196 LysGluMethrlyrsVallenaIa 203
 ||||||||||||||||||||||||
 Db 2188 AAGGAGATGACTTAAGTTCTGGCC 2211
 RESULT 48
 AF288131/c
 LOCUS AF288131 4223 bp DNA linear BCT 28-JUN-2001
 DEFINITION Nostoc punctiforme TTPA, Patn, and Pnrc genes, complete cds.
 AF288131
 ACCESSION AF288131
 VERSION AF288131.2 GI:14574706
 KEYWORDS
 SOURCE Nostoc punctiforme.
 ORGANISM Nostoc punctiforme.
 Nostoc punctiforme.
 Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 REFERENCE 1 (bases 1 to 4223)
 AUTHORS Wong,F.C.Y. and Weeks,J.C.
 TITLE A novel gene required for normal heterocyst differentiation pattern
 in Nostoc punctiforme
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4223)
 AUTHORS Wong,F.C.Y. and Weeks,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-2000) Microbiology, University of California, 1
 Shields Ave, Davis, CA 95616, USA
 REFERENCE 3 (bases 1 to 4223)
 AUTHORS Wong,F.C.Y. and Weeks,J.C.
 TITLE Direct Submission

JOURNAL Submitted (28-JUN-2001) Microbiology, University of California, 1 Shields Ave, Davis, CA 95616, USA
REMARK Sequence update by submitter
COMMENT On Jun 28, 2001 this sequence version replaced gi:9837512.
FEATURES
Source Location/Qualifiers
1..4223

CDS
/organism="Nostoc punctiforme"
/db_xref="taxon:63737"
181..750
/note="Similar to tryptophan synthase alpha-subunit"
/codon_start=1
/transl_table=1
/product="TrpA"
/protein_id="AAK68686.1"
/db_xref="GI:14574707"
/translation="MLOGITPKLRSPVLTYYPIHLRGIDKFLQETIAAGVAGLV
PDLPLEAAGLLEPAKEMGIDVLLVAPTSAPAKIEIAHSOGFYLVSVTGVR
SOLSRVSDLKLRGVTKEPIGVGFSIDMAORQYKEMGDAIIVGSAYVKRLAEG
TPBOGLSAIADLCQSLKAAIKTSTNPID"
992..1654
/function="Required for normal heterocyst differentiation"
/note="putative transmembrane protein"
/codon_start=1
/transl_table=1
/product="PatN"
/protein_id="AAG00577.1"
/db_xref="GI:9837513"
/translation="MSBSMAFTGVANAGLAALVLKGTNTPLOPNFNAVAPQMPCTV
APGIDPQMPYPTGPGYPPNPQPTATSDORLEMEKLNAAQMLERLKNQDLK
ONQOLQOVONFNTQOQMOAQNOQKVALIQPOTSWMSPLMAVAGALITIGGV
VVAQVLAFLFSRQRPARTVQVIPHQCGTPTLVVRRAEFLPASRMEARVAREDE
MH"
CDS complement(1800..3698)
/note="Similar to Arabaena PCC 7120 serine/threonine
protein kinase Pkc"
/codon_start=1
/transl_table=1
/product="Pkc"
/protein_id="AAK68687.1"
/db_xref="GI:14574708"
/translation="MIGKLDHRYQVIRVLAMGFGOTYIAQDTRRPNCIVYKHLK
PGDPRVFDAKRLFENSEAETLEKIGNHDQIPRLAYFENQEFYLVQVEYEGHTLAE
ELTPGKMSQSVQILOEVEYLEFVHSGVIRHDIKPNIIIRASDNKLVLDLFGA
VKOLRTQMTVVGQASATVVGTPTEGOGKPRENSDIYSGLIITAIQALGLLP
TALOEDPBTGIIIMQSVTVYRLAULTKVRVHFQDRYQNAIEALQACDAINPAP
TLSDPQSTSRPOVSRLOQYVAAPAMPVPRKDSKSDPILIGLIGVAAALY
GNTLPYKKNLALNFTGKDTLVNRCALVVGNSNI RSEPSINSNIIQTVADNTSE
VTGRKTRGWELEVLKSGRLAMAPAVIINNOEWASCLREKIGATQIVNDSLIATRP
TPKRVKSKDVATPLPEKLESTDDVKSEKSDPVSANIIEQAKOKVDSGDIIVGAI
ALKSVNANASGIGIOTGKMAQMODMAKAEALSNEINKAIDDGKMDKVIDYENHPE
KLPTKTYRNKIEPLFKQAAENLAKALQTLKNGNKRSTQOKLPDINQPTQENPNS
KETPKSP"

BASE COUNT 1105 a 901 c 967 g 1250 t
ORIGIN

Alignment Scores:
Pred. No.: 180 Length: 4223
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x AF288131 (1-4223)

QY 28 GluLeaSLyAlaIleAspASP 35
|||||
Db 2066 GAGATCACAAGCAATAGATCAT 2043
FCSSMONC
LOCUS FCSSMONC 4577 bp ss-RNA 1linear VRL 02-AUG-1993
DEFINITION Feline sarcoma virus (McDonough strain) transforming gene, coding

for p160gag-fms, complete coding sequence.
ACCESSION K01643
VERSION K01643.1 GI:323890
KEYWORDS p160gag-fms gene; c-myc proto-oncogene; fms oncogene; transforming gene.
SOURCE Feline sarcoma virus (McDonough strain) DNA.
ORGANISM Feline sarcoma virus
VIRUSES: Retroviral viruses
retroviruses; 1-Mammalian type C virus group.
1 (bases 1 to 4577)
Hampe,A., Gobet,M., Sherr,C.J. and Galibert,F.
Nucleotide sequence of the feline retroviral oncogene v-fms shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases
Proc. Natl. Acad. Sci. U.S.A. 81 (1), 85-89 (1984)
JOURNAL MEDLINE
PUBMED 84119469
6582485
[1] shows that a region of the v-fms gene is partially homologous to v-onc encoded tyrosine kinases.

FEATURES
Source Location/Qualifiers
1..4577
/organism="Feline sarcoma virus"
/db_xref="taxon:11772"
1..4536
/note="p160gag-fms polypeptide"
/codon_start=1
/protein_id="AAA3045.1"
/db_xref="GI:323891"
/translation="MSGASCTATGARLFGISSVIGEVRLIGDEGAGPSRSPSEVS
SVWYRSRAARVILCLVASFLVCLTELAEVGVGVTVTPSLTLDHWSVVRRAH
GIVERRKKMTLCEAEVMNNVMPREGPPPLDPTSOVERKIRIAPRHGHPDVPYI
TWSRLATDPPSWVPLPPKPPPTPLPPSPQSPAPTSLSLVLPKPPKPPPL
PPDPSPLDILTEPPPPYKCGHCPSPSGPLTPASPLASLRERRRNPESQALPL
REGPNRPQVTPFASDLVNNKLNPPSDPVALNLIESLTVTHQPTNDCCQLQ
ALLAEERQVLELARKOVPEDEGRPPQVDEAPPLRPNDFAHPAGREGLRY
RQLLAGLRGARRPNTLNAQYQVOKEEPASFLRLKAYMYTPDEBEDGOAA
SVLSFTYSSPDRLNKLQLEGLQFTLSDLKAEKLYNKRTEPEERERLROOE
ERDKRHKREMTKVATVTONRKNRDEKVGSGHGASETGPPPLCPASCCCL
PEAMGPRALLVILMATMAHQAQVPIVOPSGPELVPRGTVTVRCVNGNSVEDGPI
SPHNLDLDPSSILTTNNATFQVITGTHCTECPNPGCANATHLYIKDPRPKVLA
QEVVNLGQDLPLCLLTPALAEVGSLLVRGRPVLRQTVSFSPMHGFTHAKFL
ENHYVQCSARVDGRTVSMGIMLVQKDISAPAVLTLEPAVLVIOEAAQIVCSAN
IDVFEVDSLRIGDRTKLTISQSDQEDHRYGVQVILNLNDSFODAGVSCATYAMGN
HSASMVRRVSEASVSNLTSBDSILOETVGVKQVLOVKEVYVPLTESPMNYLIGPFD
YKDLVDVTIKDRTYRSTLSPLPKSESRSYFLARNAGONALPELITRPVRY
KVMTLLNGSOTLCEASGTFQPSVTVQCSHSDRCDSEAGVLDESHSVLSQVP
LEVTVHSLILAGTLEHNRTECPAFNSVGNSSQTFWPIISGAHPPLDEDLFTVLLT
CMSIMALLLILLLILLYKQKPKYQVWKIIESEYEGNSYTFIDPQLPYNEKEPFR
NNLOFGTLTGAGKGVYEAATAFGLKEDAVIKVAKMLSTAHADKEALMSLKT
SHLGOHNHIVLIGACVHGGVAVITVECCYGLDLNLRQAEAMPQSPSLVGDDPEA
GACVKNHILEKKYVRDSCGSSQGVDTYVEMRPVSTSSNSFSEEDLGRKEDGPLE
RDLHFSSOVAHQGAFLASKNCHIRDVAAARVLLTSGRAKIGDFGLARDIMDSNTI
VKGNARLPVKWMAPEISIFDCVYTVOSVDSGILLMEIFSLGPNPYGILLVNSKPYVL
VKDGYMAOPAFAPAKNYSIMQACMALEPRTTPTFOICSLQKQAOEDRRVPTNL
PSSSSRLRPWRGPTL"

sig_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide
BASE COUNT 1063 a 1388 c 1196 g 930 t
ORIGIN

Alignment Scores:
Pred. No.: 193 Length: 4577
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
Gaps: 14

US-10-024-955-7 (1-213) x FCSSMONC (1-4577)
QY 196 LysGluMetThrLysValIleuAla 203
|||||
DB 1546 AAGGAATGACTAAGTCTGCGC 1569
RESULT 50
L78665 5819 bp DNA linear BCF 30-DEC-1999
LOCUS L78665
DEFINITION Methylobacillus flagellatum aspartate aminotransferase (aat),
membrane protein (orf-1), homoserine dehydrogenase (hom), and
threonine synthase (thrc) thymidylate synthase (thya) genes,
complete cds.
ACCESSION L78665
VERSION L78665.1 GI:6647284
KEYWORDS aat gene; aspartate aminotransferase; hom gene; homoserine
dehydrogenase; membrane protein; thrc gene; threonine synthase;
thya gene; thymidylate synthase.
SOURCE Methylobacillus flagellatus.
ORGANISM Bacteria; Proteobacteria; beta subdivision; Methylobacillus group;
Methylobacillus.
REFERENCE 1 (bases 1 to 5819)
Marchenko,G.N., Marchenko,N.D., Tsygankov,Y.D. and
Chistoserdyov,A.Y.
TITLE Organization of threonine biosynthesis genes from the obligate
methyloctroph Methylobacillus flagellatus
JOURNAL Microbiology 145 (Pt 11), 3273-3282 (1999)
MEDLINE 2005654
PubMed 10389737
FEATURES
source
1. location/Qualifiers
1..5819
/organism="Methylobacillus flagellatus"
/strain="Kt"
/db_xref="taxon:405"
/complement(20..388)
/gene="orf-1"
/complement(20..388)
/gene="orf-1"
/note="predicted amino acid sequence of the orf-1 gene has
homology to 13 kD major membrane protein of Francisella
tularensis and 12.7 kD protein of Rhodobacter capsulatus.;
hypothetical; putative"
/codon_start=1
/transl_table=11
/product="membrane protein"
/protein_id="AAF21127.1"
/db_xref="GI:6647285"
/translation="MKLHLTQAAGSHLITGYGDGWEINHQRHESLIIVPGSILLPM
PYEDFDLOPTHEOALALPKPELLILGTGSRHRLPSPISGRDLVVGISMECWTAA
CRPTNYIMAESQVAALIT"
499..1788
/gene="aat"
499..1788
/gene="aat"
/FC_number="2.6.1.1"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="aspartate aminotransferase"
/protein_id="AAF21128.1"
/db_xref="GI:6647286"
/translation="MQEISKSKLSNWVCYDIRPVLDIPRAROMEEDGRHIIKINQMA
PREFAPPEILQDVIRNMDASGYTDSKGLPAARKAIMHTQCKNIQGYTIDITIGN
GVSELTVMQAQALINNDQVLYPMDYPLMTAAVNLAGGTARUYVDEQGTWLPDLRD
IENKTIANTRGIVILINPNPTGALYPRELEGLIETARHGLVTFADETDYKVLPGN
THTSIASLADVLEVFENGLSKNRYTCGYRAGLVISGGRHAGDYIEGLMLASMR
LNRQCARAVGDDPRAGVOSIDLVARTGLRCORDLAYKMLTDIPVSCVKPAGLYL
FRLDPKTYPIODDOOFTIDILLEKVLIVCGGFPMWRAPDHRVYFLFNVDDLTFAP
TRARPLENPKRKHSTITVEAKPRKRAQTSSASA"
1840..3153
gene

CDS
/gene="hom"
1840..3153
/gene="hom"
/FC_number="1.1.1.3"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="homoserine dehydrogenase"
/protein_id="AAF21129.1"
/db_xref="GI:6647287"
/translation="MKPIWGLLIGITGVGGTYVTLTRNOEETARRAGRPPIATRVAD
RULELARKVTAGOVVTDPAFVAVADPSIDIVELYIGCTVARELVMAIEKHVHT
ANKGILALNGNEIIPAAQOKGYIVAFEAIVAGCPIIRAKVRGLANREWIAGIING
TTFNFIUSEMERKALAFADYLAQRDLGAEDPTDEBSVDAAHKMLMILIAIGGIPM
OPDKAVEGISKLSIDIRYAEELGRVRLILITRKTSOGVELRVHPPLIPERLIAN
VNGMANAVLVKGDVAGPPLYGAGAGAEPTASAVADIVDYVTRHTTVDHQRVPLAF
OPDOLVDPLILAIGEVSSAYYLRRVRADVDPGLVADVTRILGDRQISIDAMIQKEPOEG
EEOADIIITHTYTAENMDALAIETALPAISGSVRLMEELSR"
3204..4634
/gene="thrc"
3204..4634
/gene="thrc"
/FC_number="4.2.99.2"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="threonine synthase"
/protein_id="AAF21130.1"
/db_xref="GI:6647288"
/translation="MRYISTRGQSPRQSTTEILLGLAFDGLIYPEQYPOISQDELN
AMRGMYRDLAFTIILSLIDIPAAADIRALIIDTYRADVIYARAGQNAEDITPTL
KLEDDLYLISLNGPTLAFKPMADLGNLFPEYMLAOKGTINILGATISGDGSAEY
AMRGKGVVFMILSPYOKMSRFQTAQIYSLQDNINILAVKGFDDCODIVKAVSNDH
AFKAKKICAVANSINMARVAAOVVYFYFKGYFAVTDNAORVSEAVSPSGNFGVACGHI
ARMGILPIAKIVLATNENDVLDPEFKTGYRRGASANTYHSSPMDSKASNFERY
FDLVGRYIAKVRKELMKRYDAGGSVDLNEGSGWAKYADGVFVSGSSNHANGKOTTYTH
ERYGVITDTHADGLKVALQRRAGIPMLVLTETALPAKEDAIVEALIGOKPARATMR
LQSLPQFEVMEADVAALIKQFIVEHV"
4647..5441
/gene="thya"
4647..5441
/gene="thya"
/FC_number="2.1.1.45"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="thymidylate synthase"
/protein_id="AAF21131.1"
/db_xref="GI:6647289"
/translation="MKVYHDLKRVHLEQGHKKEDRTGTGTLVSFGYQMRFDLSEGFPL
LTKTKVHLNAIILHELIMFLOGSTNIAYLKENGVTITWDEWADEGNIGPVYQWRMP
KPDGHIIDISFVIAIKSNPSRLIIVSAMNVADVDEKMLPCCHALFOFYVAEKLIS
COLYORSADIFIGVNPENIASYVALITPMVAVRQVDLELGDVFLGDAHIYLNLEQYKE
QLSREPRYLPVNRINDEVDIFAFREDTTLNYSHPAIKAPVAV"
BASE COUNT 1244 a 1609 c 1712 g 1254 t
ORIGIN
Alignment Scores:
Pred. No.: 239 Length: 5819
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 1 Gaps: 0
US-10-024-955-7 (1-213) x L78665 (1-5819)
QY 34 AspAspAlaIleAlaIleAlaIleAla 41
|||||
DB 3070 GATGACGCGATTCGCCGATCGAG 3093
Search completed: February 20, 2003, 19:07:10

Fri Feb 21 10:04:00 2003

us-10-024-955-7.rge

Page 31

Job time : 2430 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 18:08:31 ; Search time 268 Seconds
(without alignments)
1789.834 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213

Sequence: 1 MKKFLIAAFAVAVASADPL.....VRKEMTKVLAPAFKRELEKN 213

Scoring table:

OLIGO	Xgapop 60.0	Xgapext 60.0
	Ygapop 60.0	Ygapext 60.0
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4368707

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:
-MODL=frame_p2n.model -DEV=xlp
-O=/cgn2.1/USFPO/spool/US10024955/unat_14022003.130735.24213/app.query.fasta.1.391
-DB=N_Geneseq_101002 -QFMT=fastap -SUFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -STAR=1 -END=1 -MATRIX=0.190 -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=50 -MODE=LOCAL
-OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10024955.ecgn.1.1.396.&runat_14022003.130735.24213 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONLOG -DEV-TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXP=7

Database : N_Geneseq_101002.*

1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213	100.0	761	15	AAO71401
2	25	11.7	812	15	AAO71400
3	9	4.2	317	24	ABK39008
4	8	3.8	24	15	AAO71405
5	8	3.8	261	24	ABN68278
6	8	3.8	673	21	AAC08384
7	8	3.8	745	23	ABL03859
8	8	3.8	1050	21	AAC42988
9	8	3.8	1050	21	ABN66096
10	8	3.8	1128	22	AAH68440
11	8	3.8	1239	14	AAO48930
12	8	3.8	1833	22	AAH27069
13	8	3.8	2834	23	ABL25718
14	8	3.8	2917	23	ABL03858
15	8	3.8	3023	19	AAV49262
16	8	3.8	3050	23	ABL16976
17	8	3.8	3454	23	ABL02294
18	8	3.8	3458	23	ABL02292
19	8	3.8	3674	13	AAO29114
20	8	3.8	6027	20	AAI13061
21	8	3.8	7318	23	ABL04928
22	8	3.8	8440	17	AAT13265
23	8	3.8	11623	22	AAK81402
24	8	3.8	19983	24	AAD32835
25	8	3.8	20000	24	AAD32823
26	8	3.8	20000	24	AAD32837
27	8	3.8	349980	22	AAH68530
28	7	3.3	30	18	AAT90312
29	7	3.3	30	20	AAK76935
30	7	3.3	211	21	AAA41893
31	7	3.3	222	15	AAO73773
32	7	3.3	261	24	ABN68279
33	7	3.3	264	23	AAS55545
34	7	3.3	289	24	ABL73601
35	7	3.3	290	22	AAK78516
36	7	3.3	294	20	AAV86640
37	7	3.3	306	24	ABK74171
38	7	3.3	334	18	AAV77349
39	7	3.3	401	22	AAK95608
40	7	3.3	401	22	AAK97101
41	7	3.3	408	22	AAT90301
42	7	3.3	419	22	AAK62472
43	7	3.3	446	23	ABV15457
44	7	3.3	471	24	ABL83546
45	7	3.3	489	23	ABL13765

ALIGNMENTS

RESULT 1
AAO71401
ID AAO71401 standard; CDNA: 761 BP.

AAO71401:
AC
XX
AC
XX
DT
XX
DE
XX
KW
XX
OS
XX
FT
XX
FT
XX

01-APR-1995 (first entry)

House dust mite allergen DerfVII CDNA.

DerfVII allergen; antiallergic; allergy diagnosis; ss.

Dermatophagoides farinae.

key
FH
FT
FT

Location/Qualifiers
43..681
/*tag a

misc_feature 94..111

```

FT      /*tag= b
FT      /note= "Df1 primer"
FT      247..263
FT      misc_feature
FT      /*tag= c
FT      /note= "Df4 primer"
FT      559..582
FT      misc_feature
FT      /*tag= d
FT      /note= "Df5 primer"
FT      684..664
FT      misc_feature
FT      /*tag= e
FT      /note= "Df6 primer"
FT      225..208
FT      misc_feature
FT      /*tag= f
FT      /note= "Df8 primer"
FT      WO9420614-A.
XX
PD      15-SEP-1994.
XX
PF      11-MAR-1994; 94MO-AU00117.
XX
PR      12-MAR-1993; 93US-0031141.
PR      22-JUN-1993; 93US-0081540.
XX
PA      (CHIL-) INST CHILD HEALTH RES.
XX
PI      Chua K, Thomas WR;
XX
DR      WPI: 1994-303021/37.
DR      P-PSDB; AAR60576.
XX
PT      New nucleic acid encoding specific dust mite allergens - and
PT      related vectors, transformed cells, peptides and antibodies,
PT      useful for desensitisation and diagnosis.
XX
PS      Claim 13; Page 38-39; 67pp; English.
XX
CC      DerpVII antigen is useful as antiallergic reagent for treating
CC      sensitivity to house dust mite allergens. The DNA can be used
CC      as a probe to detect the sensitivity of an individual to the
CC      allergen.
XX
SQ      Sequence 761 BP; 245 A; 134 C; 137 G; 245 T; 0 other;

Alignment Scores:
Pred. No.: 1.95e-211 Length: 761
Score: 213.00 Matches: 213
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
15

US-10-024-955-7 (1-213) x AAQ71401 (1-761)
QY      1 MetMetLysPheLeuLeuIleAlaAlaValAlaPheValAlaSerAlaAspProIle 20
DB      43 ATGATGAAATTTTGTGATTGCTGCCGTGCAATTGTGCGCTTGCGCTGATCCAATT 102
QY      21 HistYAspIysIlePhrGluGluIleAsnIysAlaIleAspAspAlaIleAlaIle 40
DB      103 CACTATGATTAATACCGCAGAAATCAACAAAGCTATTGATGATGCGCATTTGCTATT 162
QY      41 GluGlnSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGluArg 60
DB      163 GAACATCCGAAACAAATAGATCCATGAAGTACCTGATCATGCCGATTAATCGAAGCT 222
QY      61 HisValGlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGly 80
DB      223 CATTGTGGTATTGTGATTCAAAAGTGATGATGATGATGATGATGATGATGATGATGAT 282
QY      81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys 100
DB      283 TTGAACCAATGAAACGCTCAAGGTGATGATTAATGTCAAAAGGTGAAGGGTATTGTTAAA 342

```

```

QY      101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluThrAspLeuAlaTyr 120
DB      343 GCTCATTTGTTGATCGGTGTTCACGATGATATGCTGTCGATGGAATATGATTAGCATAC 402
QY      121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValAla 140
DB      403 AAATGGGTGATCTTCATCCAACCACTCATGTCATTTCCGATATTCAGATTTTGTGTTGT 462
QY      141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetPheSerPheGluValArg 160
DB      463 GCCTTGCCCTGTAATTTCTGATGAAGGTAAACATTAACATGACATCTTTGAAGTACGA 522
QY      161 GlnPheAlaAsnValAlaValAsnHisIleGlyLysLeuSerIleLeuAspProIlePheGly 180
DB      523 CAATTCGCTAATGTGTGCACCAACATATTGCGTCTTTCATATCTTGATCCAAATTTTGGG 582
QY      181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys 200
DB      583 GTTTTATCTGATGATATTGACCGCTATTTCAGACACCGCTACGTAAGCAAAATGACCAAA 642
QY      201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
DB      643 GATTGGCACCCACCATTTTAACGTGAATTGGAAAAAAT 681

RESULT 2
AAQ71400
ID      AAQ71400 standard; cDNA; 812 BP.
XX
AC      AAQ71400;
XX
DT      01-APR-1995 (first entry)
XX
DE      House dust mite allergen DerpVII cDNA.
XX
KW      DerpVII allergen; antiallergic; allergy diagnosis; ss.
XX
OS      Dermatophagoides pteronyssinus.
XX
FH      Key Location/Qualifiers
FT      CDS 68..712
FT      /*tag= a
FT      119..136
FT      misc_feature
FT      /*tag= b
FT      /label= primer
FT      /note= "SEQ ID NO:3"
FT      misc_feature
FT      272..288
FT      /*tag= c
FT      /label= primer
FT      /note= "SEQ ID NO:4"
FT      misc_feature
FT      584..607
FT      /*tag= d
FT      /label= primer
FT      /note= "SEQ ID NO:5"
FT      FT

WO9420614-A.
PD      15-SEP-1994.
XX
PF      11-MAR-1994; 94MO-AU00117.
XX
PR      12-MAR-1993; 93US-0031141.
PR      22-JUN-1993; 93US-0081540.
XX
PA      (CHIL-) INST CHILD HEALTH RES.
XX
PI      Chua K, Thomas WR;
XX
DR      WPI: 1994-303021/37.
DR      P-PSDB; AAR60575.
XX
PT      New nucleic acid encoding specific dust mite allergens - and
PT      related vectors, transformed cells, peptides and antibodies,

```

PT .useful for desensitisation and diagnosis.
XX
PS Claim 3; Page 35-36; 67pp; English.
XX
CC DerpVII antigen is useful as antiallergic reagent for treating
CC sensitivity to house dust mite allergens. The DNA can be used
CC as a probe to detect the sensitivity of an individual to the
CC allergen.
XX
SQ Sequence 812 BP; 270 A; 133 C; 142 G; 267 T; 0 other;
Alignment Scores:
Pred. No.: 2,08e-16 Length: 812
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.74% Indels: 0
DB: 15 Gaps: 0
US-10-024-955-7 (1-213) x AAQ71400 (1-812)
QY 155 ThrSerPheGluValArgGlnPheAlaAsnValAlaHisIleGlyGlyLeuSerIle 174
DB 530 KCATGTTGGAAGTACGATTCATTCGCCATGTTGTCATCATATTCGTCCTTTCATTT 589
QY 175 LeuAspProIlePhe 179
DB 590 TTGGATCCATTTTC 604
RESULT 3
ABK39008/c
ID ABK39008 standard; cDNA: 317 BP.
XX
AC ABK39008;
XX
DT 21-MAY-2002 (first entry)
XX
DE cDNA encoding lung tumour protein clone R0128:E12.
XX
KW Lung tumour; cancer; T cell; immune response stimulator;
KW cytosolic; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200204514-A2.
XX
PD 17-JAN-2002.
XX
PF 10-JUL-2001; 2001MO-US22058.
XX
PR 11-JUL-2000; 2000US-0614124.
PR 29-AUG-2000; 2000US-0651563.
PR 08-SEP-2000; 2000US-0658824.
PR 26-SEP-2000; 2000US-0671325.
PR 06-OCT-2000; 2000US-0677419.
PR 30-OCT-2000; 2000US-0702705.
PR 13-DEC-2000; 2000US-0736457.
PR 03-MAY-2001; 2001US-0849626.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Reiter MM;
PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
PI McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;
XX
DR WPI: 2002-164634/21.
XX
PT Novel polynucleotide encoding a lung tumour polypeptide useful for
PT stimulating and/or expanding T cells specific for a tumour protein -
XX
PS Claim 1; SEQ ID No 1046; 223pp; English.
XX
CC The invention describes an isolated polynucleotide and polypeptide

CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein for determining the presence of a cancer in a patient. A
CC composition containing the polynucleotide and/or polypeptide is useful
CC for treating a lung cancer in a patient. The polypeptide is useful for
CC removing tumour cells from a biological sample. The polynucleotide is
CC also useful as probe or primer to detect the level of mRNA encoding a
CC tumour protein. This sequence encodes a lung tumour associated protein
CC or protein fragment, described in the method of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 317 BP; 57 A; 84 C; 110 G; 66 T; 0 other;
Alignment Scores:
Pred. No.: 3.37 Length: 317
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 24 Gaps: 0
US-10-024-955-7 (1-213) x ABK39008 (1-317)
QY 188 AlaIlePheGlnAspThrValArgLys 196
DB 309 GCCATTTCACGACACGCGTGAAG 283
RESULT 4
AAO71405
ID AAO71405 standard; cDNA: 24 BP.
XX
AC AAO71405;
XX
DT 01-APR-1995 (first entry)
XX
DE Primer Df5 for house dust mite allergen DerfVII cDNA.
XX
KW Primer: Df5; DNA sequencing; DerfVII allergen; antiallergic;
KW allergy diagnosis; ss.
XX
OS Dermatophagoides farinae.
XX
PN WO9420614-A.
XX
PD 15-SEP-1994.
XX
PF 11-MAR-1994; 94WO-AU00117.
XX
PR 12-MAR-1993; 93US-0031141.
PR 22-JUN-1993; 93US-0081340.
XX
PA (CHIL-) INST CHILD HEALTH RES.
XX
PI Chua K, Thomas WR;
XX
DR WPI: 1994-303021/37.
XX
PT New nucleic acid encoding specific dust mite allergens, - and
PT related vectors, transformed cells, peptides and antibodies,
PT useful for desensitisation and diagnosis.
XX
PS Example 5; Page 33; 67pp; English.
XX
CC The DNA sequencing primer Df5 is derived from nucleotides 559-582
CC of DerfVII (AAO71401) and is used in the polymerase chain reaction
CC amplification and sequencing of a cDNA clone encoding DerfVII from a
CC phage lambda-gt10 cDNA library. DerfVII antigen is useful as an
CC antiallergic reagent for treating sensitivity to house dust mite
CC allergens.
XX
SQ Sequence 24 BP; 5 A; 5 C; 4 G; 10 T; 0 other;

	Alignment Scores:	
Pred. No.:	3..2	Length:
Score:	8..00	Matches:
Percent Similarity:	100.00%	Conservative:
Best Local Similarity:	100.00%	Mismatches:
Query Match:	3..76%	Indels:
DB:	15	Gaps:
		0

US-10-024-955-7 (1-213) x AAQ71405 (1-24)	
OY	173 SerfLeuaspProilepegly 180
Db	1 TCAATCTTGATCCAAATTTTGGC 24

RESULT 5	
ABN68278	
ID	ABN68278 standard; DNA; 261 BP.
XX	
AC	ABN68278;
XX	
DT	01-JUL-2002 (first entry)
XX	
DE	Streptococcus polynucleotide SEQ ID NO 4469.
XX	
KW	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW	group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW	antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS	Streptococcus agalactiae.
XX	
PN	WO200234771-A2.
XX	
PD	02-MAY-2002.
XX	
PF	29-OCT-2001; 2001MO-GB04789.
XX	
PR	27-OCT-2000; 2000GB-0026333.
PR	24-NOV-2000; 2000GB-0028727.
PR	07-MAR-2001; 2001GB-0005640.
XX	
PA	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
PI	Tetteijn H;
XX	
DR	WPJ: 2002-352536/38.
XX	
DR	P-Psdb: ABP27647.
XX	

New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein -

Claim 7; Page 3608; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (II), ABN66044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.

XX	Sequence	261 BP;	84 A;	48 C;	59 G;	70 T;	0 other;
50	Alignment Scores:						
	Pred. No.:	30.6	Length:	261			
	Score:	8.00	Matches:	8			
	Percent Similarity:	100.00%	Conservative:	0			
	Best local Similarity:	100.00%	Mismatches:	0			
	Query Match:	3.76%	Indels:	0			
	DB:	24	Gaps:	0			
	US-10-024-955-7 (1-213) x ABN68278 (1-261)						
OY	34 AspAspAlaIleAlaIleAlaIleGlu 41						
Db	211 GATGATGCATTGCACGCTATCGAA 234						
RESULT 6							
AAC08384/c							
ID	AAC08384 standard; cDNA; 673 BP.						
AC	AAC08384;						
XX							
DT	06-OCT-2000 (first entry)						
XX							
DE	Human secreted protein 5' EST, SEQ ID NO: 12459.						
XX							
XX	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;						
KW	gene therapy; chromosome mapping; ss.						
XX							
OS	Homo sapiens.						
XX							
PN	EP1033401-A2.						
XX							
PD	06-SEP-2000.						
XX							
PF	21-FEB-2000; 2000EP-0200610.						
XX							
PR	26-FEB-1999; 99US-0122487.						
XX							
PA	(GEST) GENSET.						
XX							
PI	Dumas Milne Edwards J, Duclert A, Giordano J;						
XX							
DR	WPI: 2000-500381/45.						
XX							
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for						
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for						
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -						
XX							
PS	Claim 1; SEQ ID 12459; 71pp + CD-ROM; English.						
XX							
CC	The present sequence is one of a large number of 5' ESTs derived from						
CC	mRNAs encoding secreted proteins. No ORF has yet been conclusively						
CC	identified within the present sequence. The 5' ESTs were prepared from						
CC	total human RNAs or poly(A) RNAs derived from 30 different tissues. EST						
CC	sequences usually correspond mainly to the 3' untranslated region (3'UTR)						
CC	of the mRNA because they are often obtained from oligo-dT primed cDNA						
CC	libraries. Such ESTs are not well suited for isolating cDNA sequences						
CC	derived from the 5' ends of mRNAs and even in those cases where longer						
CC	cDNA sequences have been obtained, the full 5' UTR is rarely included.						
CC	5' ESTs are derived from mRNAs with intact 5' ends and can therefore be						
CC	used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used						
CC	in diagnostic, forensic, gene therapy and chromosome mapping procedures.						
CC	They are used to obtain upstream regulatory sequences and to design						
CC	expression and secretion vectors.						
XX							
S0	Sequence 673 BP; 222 A; 122 C; 112 G; 215 T; 2 other;						
Alignment Scores:							
Pred. No.:	74.9	Length:	673				
Score:	8.00	Matches:	8				
Percent Similarity:	100.00%	Conservative:	0				

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 21 Gaps: 0

US-10-024-955-7 (1-213) x AAC08384 (1-673)

QY 167 AsnHisIleGlyLeuSerIle 174
|||||
DB 337 ATACATTGCTGTTGACGATA 314

RESULT 7

ABL03859/c
ID ABL03859 standard; cDNA: 745 BP.

AC ABL03859;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6059.

KW Drosophila: developmental biology; cell signalling; insecticide;
pharmaceutical; gene: ss.

OS Drosophila melanogaster.

XX MO200171042-A2.

PN 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR P-PSDB: ABB59756.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

PS Claim 1; SEQ ID NO 6059; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
sequences (ABB57737-ABB72072).
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 745 BP; 224 A; 172 C; 188 G; 161 T; 0 other;

Alignment Scores:

Pred. No.: 82.4 Length: 745
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 23 Gaps: 0

US-10-024-955-7 (1-213) x ABL03859 (1-745)

QY 8 AlaAlaValAlaPheValAlaVal 15
|||||

DB 677 GCTGCTGTTCCTTTGTTGCTGTT 654

RESULT 8

AAC42988
ID AAC42988 standard; DNA: 1050 BP.

AC AAC42988;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37607.

XX Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136382.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

RESULT 9
ABN66096/c
ID ABN66096 standard; DNA: 1050 BP.
XX
XX
AC ABN66096;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 105.
XX
KM Streptococcus; GAS; GB5; group B streptococcus; Streptococcus agalactiae;
KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX
PA (CHTR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Maignani V, Margalit Ros YI, Grandi G, Fraser C;
PI Tettein H;
XX
XX
DR WPI: 2002-352536/38.
DR P-PSDB; ABP25465.
XX
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX
PS Claim 7; Page 3164; 4525bp; English.
XX
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 1050 BP; 301 A; 195 C; 248 G; 306 T; 0 other;

Alignment Scores:
Pred. No.: 114 Length: 1050
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 24 Gaps: 0

US-10-024-955-7 (1-213) x ABN66096 (1-1050)
QY 115 glutyraspleualatylrlyslen 122
|||||

DB 682 GAGTACAGCTGCATATAGCTA 659
RESULT 10
AAH68440/c
ID AAH68440 standard; DNA: 1128 BP.
XX
XX
AC AAH68440;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 3475.
XX
KM Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX
DR WPI: 2001-376931/40.
DR P-PSDB; AAG93221.
XX
XX
PT Novel polynucleotides derived from Corynebacterium, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
PS Claim 1; SEQ ID NO: 3475; 246bp + Sequence Listing; English.
XX
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1128 BP; 286 A; 292 C; 312 G; 238 T; 0 other;

Alignment Scores:
Pred. No.: 122 Length: 1128
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 22 Gaps: 0

US-10-024-955-7 (1-213) x AAH68440 (1-1128)
QY 8 AlaAlaValAlaIaphValAlaVal 15
|||||
DB 38 GCGGAGTTGCTTGTCTGTC 15
RESULT 11

```

AAQ48930
ID AAQ48930 standard; DNA; 1239 BP.
XX
XX AAQ48930;
AC
XX
XX 23-MAR-1994 (first entry)
DE Methylobacillus glycyogenes 1006 homoserine dehydrogenase.
XX
XX L-threonine biosynthesis; homoserine dehydrogenase;
XX threonine synthase; homoserine kinase; amino acid; ATCC 21371; ss.
XX
XX Methylobacillus glycyogenes strain 1006 (ATCC 21371).
OS
XX
XX Key Location/Qualifiers
XX CDS 1..1239
XX /tag=a
XX /product= homoserine_dehydrogenase
XX
XX JP0520786-A.
XX
XX 20-AUG-1993.
XX
XX 29-JAN-1992; 92JP-0014335.
XX
XX 29-JAN-1992; 92JP-0014335.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX WPI; 1993-297465/38.
XX
XX P-PSDB; AAR41823.
XX
XX Prep. of L-threonine by fermentation in methanol soln. - using
XX Methylobacillus sp. conf. recombinant vector to produce
XX homoserine-dehydrogenase or kinase, or threonine synthase
XX
XX Claim 5; Page 14-15; 21pp; Japanese.
XX
XX DNA coding for at least one of the enzymes homoserine dehydrogenase,
XX threonine synthase and homoserine kinase is incorporated into a
XX recombinant expression vector and used to transform Methylobacillus
XX bacteria. The transformants are cultured in medium containing
XX methanol as carbon source to synthesize L-threonine.
XX See also AAQ48929 and AAQ48931-Q48932.
XX
XX SQ Sequence 1239 BP; 272 A; 318 C; 351 G; 298 T; 0 other;

Alignment Scores:
Pred. No.: 133 Length: 1239
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 14 Gaps: 0

US-10-024-955-7 (1-213) x AAQ48930 (1-1239)
OY 34 AspAspAlaIleAlaIleAlaIleGlu 41
Db 1156 GATGATGCATCGCCGCCATTGAG 1179

RESULT 12
AAH27069
ID AAH27069 standard; cDNA; 1833 BP.
XX
XX AAH27069;
AC
XX
XX 21-DEC-2001 (first entry)
DE Feline leukemia virus p27-gp70 nucleic acid nFeLVp27-gp701833.
XX
XX nFeLV; glycoprotein; gp70; p27 protein; nFeLVp27-gp701833; antigen;
XX immune status; vaccination status; cat; ss.
XX

```

```

XX
XX OS Feline leukemia virus.
XX OS Synthetic.
XX
XX WO200166568-A2.
XX
XX 13-SEP-2001.
XX
XX 07-MAR-2001; 2001WO-US07251.
XX
XX 09-MAR-2000; 2000US-0521738.
XX
XX (HESK-) HESKA CORP.
XX (COLS ) UNIV COLORADO STATE RES FOUND.
XX
XX Jensen WA, Lappin MR, Rosen DK, Andrews JS;
XX
XX WPI; 2001-639000/73.
XX
XX P-PSDB; AAM50122.
XX
XX Determining immune status or vaccination status of an animal to e.g.
XX calicivirus comprises using a recombinant viral antigen -
XX
XX Claim 33; Page 116-119; 132pp; English.
XX
XX The present nucleic acid sequence, designated nFeLVp27-gp701833,
XX encodes a fusion protein (see AAM50122), termed pFeLVp27-gp70611,
XX comprising the C-terminus of feline leukemia virus (FeLV)
XX pFe5-gag and glycoprotein gp70. The nucleic acid was produced
XX by PCR amplification and TA cloning, and ligated to recombinant
XX vector lambdaDAPCrio/T2ori/RSET-B/Hisless. The recombinant molecule
XX was used to transform Escherichia coli. Transformed cells produced
XX pFeLVp27-gp70611 without an N-terminal His tag. pFeLVp27-gp70611 is
XX an example of a recombinant infectious agent antigen that can be used
XX in the method of the invention to determine the immune status of an
XX animal. The method involves contacting a biological specimen of an
XX animal (cat, dog or horse) with a recombinant antigen, and detecting
XX the presence or absence of a complex between the recombinant
XX and an antibody present in the sample. The method determines
XX whether the animal is protected against disease or should be
XX vaccinated. Recombinant antigens (see AAM50107-24), nucleic acids
XX encoding them (see AAH27054-71), methods of producing them, and
XX assay methods are provided.
XX
XX SQ Sequence 1833 BP; 551 A; 479 C; 430 G; 373 T; 0 other;

Alignment Scores:
Pred. No.: 193 Length: 1833
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 22 Gaps: 0

US-10-024-955-7 (1-213) x AAH27069 (1-1833)
OY 196 LysGluMetThrIysValIleuAla 203
Db 391 AAGGACATGACTAAAGTTCTGCC 414

RESULT 13
ABL25718/c
ID ABL25718 standard; DNA; 2834 BP.
XX
XX ABL25718;
AC
XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster genomic polynucleotide seq ID NO 28627.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX

```

OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI WPI: 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1: SEQ ID NO 28627; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 2834 BP; 789 A; 632 C; 635 G; 778 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 292 Length: 2834
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: Gaps: 0
US-10-024-955-7 (1-213) x ABLI25718 (1-2834)
OY 5 LeuLeuIleAlaIaValaIaIaPhe 12
DB 1430 CTTTAAATTGACGCTGACCTTT 1407
RESULT 14
ABLI03858
ID ABLI03858 standard; cDNA; 2917 BP.
XX
XX ABLI03858;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 6056.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PR
XX
XX 23-MAR-2000; 2000US-191637P.
PR

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI WPI: 2001-656860/75.
XX
XX P-PSDB: ABB59755.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1: SEQ ID NO 6056; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 2917 BP; 824 A; 579 C; 585 G; 929 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 300 Length: 2917
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: Gaps: 0
US-10-024-955-7 (1-213) x ABLI03858 (1-2917)
OY 8 AlaIaValaIaIaPheValaIaVal 15
DB 1069 GCTGCTGTGCTTGTCTGCTT 1092
RESULT 15
AAV49262
ID AAV49262 standard; DNA; 3023 BP.
XX
XX AAV49262;
XX
XX 28-OCT-1998 (first entry)
XX
XX FeLV-B gag/pol gene.
DE
XX
XX Multivalent vaccine; cat; pathogen; respiratory disease; FeLV; FCV;
KW digestive disease; feline leukaemia virus; feline panleukopenia virus;
KW feline calicivirus; feline immunodeficiency virus; FIV; rabies virus;
KW vector; envelope glycoprotein; ss.
XX
XX Feline leukemia virus.
OS
XX
XX Key Location/Qualifiers
FH 1.1743
FT CDS
FT
FT /tag= a
FT /product= "gag protein"
FT 1736..3023
FT /tag= b
FT /product= "pol protein"
FT /note= "the coding sequence for this protein does
FT not contain a start codon at the 5' end of
FT the sequence; the pol protein is generated
FT by a translational frameshift at the 3' end
FT of the gag gene; also the 3' end of the gene
FT contains no stop codon"

```

XX FR2/51223-A1.
PN
XX
XX 23-JAN-1998.
PD
XX
XX 19-JUL-1996; 96FR-0009337.
PF
XX
XX 19-JUL-1996; 96FR-0009337.
PR
XX
XX (INMR ) RHONE MERIEUX SA.
PA
XX
PI Audonnet JCF, Baudu P, Bouchardon A, Riviere MEA;
XX
XX WPI; 1998-112823/11.
DR
XX P-PSDB; AAM68402, AAM68403.
XX
XX Multi-valent polynucleotide vaccines against feline pathogens -
PT consist of at least 3 plasmids able to express protective antigens
PT from specified viruses
XX
XX Example 9; Fig 5; 42pp; French.
XX
XX The invention relates to a multivalent vaccine for protecting cats
XX against several pathogens, especially pathogens associated with
XX respiratory and digestive diseases. The pathogens are especially
XX selected from feline leukaemia virus (FeLV), feline panleukopenia
XX virus (FPV), feline calicivirus (FCV), feline immunodeficiency virus
XX (FIV), or rabies virus. The vaccines are preferably composed of
XX polynucleotide sequences encoding 3 antigens, all as part of vectors.
XX This sequence represents the coding region for the gag/pol proteins
XX (gag/pol) from the feline leukaemia virus (FeLV) sub-type A strain
XX Glasgow-1. The sequence was sub-cloned into the plasmid pVR1012 to
XX generate plasmid pPB181 for use in the vaccine.
XX
SQ Sequence 3023 BP; 824 A; 831 C; 714 G; 654 T; 0 other;

Alignment Scores:
Pred. No.: 310 Length: 3023
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 19 Gaps: 0

US-10-024-955-7 (1-213) x AAV49262 (1-3023)
QY 196 Lysglumethrlysvallena1a 203
Db 1537 AAGGAGTGAAGTCTGCGC 1560

RESULT 16
ABLI6976/c
ID ABLI6976 standard; DNA: 3050 BP.
XX
XX ABLI6976;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 2401.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.

```

```

XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 2401; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI840-ABLI175) and the encoded proteins
XX (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3050 BP; 925 A; 602 C; 633 G; 890 T; 0 other;

Alignment Scores:
Pred. No.: 313 Length: 3050
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 23 Gaps: 0

US-10-024-955-7 (1-213) x ABLI6976 (1-3050)
QY 8 ALAAlaValAlaPheValAlaVal 15
Db 2603 GCTGCTGTGCTTGTGTGCTGTT 2580

RESULT 17
ABLI02294/c
ID ABLI02294 standard; CDNA: 3454 BP.
XX
XX ABLI02294;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 1364.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

```

PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 1364; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 3454 BP; 916 A; 843 C; 785 G; 910 T; 0 other.

Alignment Scores:
Pred. No.: 352 Length: 3454
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 23 Gaps: 0

US-10-024-955-7 (1-213) x ABL02294 (1-3454)

OY 5 LeuLeu11eAla1aVala1aPhe 12
ID |||||||
DB 3425 CTTTAAATTCAGCTGTAGCTTTT 3402

RESULT 18
ABL02292/c
ID ABL02292 standard; cDNA; 3458 BP.
XX
AC ABL02292;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1358.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN MO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB58189.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 1358; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 3458 BP; 917 A; 844 C; 786 G; 911 T; 0 other.

Alignment Scores:
Pred. No.: 352 Length: 3458
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 23 Gaps: 0

US-10-024-955-7 (1-213) x ABL02292 (1-3458)

OY 5 LeuLeu11eAla1aVala1aPhe 12
ID |||||||
DB 3424 CTTTAAATTCAGCTGTAGCTTTT 3401

RESULT 19
AAQ29114
ID AAQ29114 standard; DNA; 3674 BP.
XX
AC AAQ29114;
XX
DT 24-FEB-1993 (first entry)
XX
DE FelV-A gag/pol fragment.
XX
XX
XX ALVAC; feline leukemia virus; FelV-A; gag; pol; pFGA-2 gag; PCR;
KW polymerase chain reaction; vaccinia virus; H6 promoter; PC3FGAG;
KW PC3FGAGVQ; expression cassette; PC3DOFGAGVQ; VCP83; VCP87; ATI;
KW rescue virus; VCP97; immunosuppressive region; VCP93; polylinker;
KW A type inclusion body region; PSD541; ss.
XX
OS Synthetic.
XX
XX
XX Key Location/Qualifiers
FH misc_feature 652..654
FT /*tag= a
FT /label= Initiation_codon
XX
PN MO9215672-A.
XX
PD 17-SEP-1992.
XX
PE 09-MAR-1992; 92WO-US01906.
XX
PR 07-MAR-1991; 91US-0666056.
PR 11-JUN-1991; 91US-0713967.
PR 06-MAR-1992; 92US-0847951.
XX
PA (VIRO-) VIROGENETICS CORP.
XX
XX Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP;
PI Limbach KJ, Norton EK, Paoletti E, Perkins ME, Pincus SE;
PI Riviere M, Tartaglia J, Taylor J;
XX
DR WPI; 1992-331718/40.
XX
PT Vaccinating against recombinant, attenuated pox-virus - use for
PT vaccinating against viral infections such as rabies, hepatitis B,
PT HIV, HSV, EBV, CMV, mumps etc.
XX
PS Disclosure; Fig 27; 456pp; English.
XX
XX The sequence given encodes the feline leukemia virus (FelV-A) gag

CC gene. This sequence was used in the construction of an ALVAC-based
CC recombinant containing the Felv-A gag gene. The gag/pol sequence was
CC derived from plasmid pFCA-2 gag. The gag gene was liberated on a 2.5
CC kb fragment and was amplified by polymerase chain reaction (PCR).
CC This sequence was aligned with the vaccinia virus H6 promoter. The
CC Felv-A pol gene was also included in this construction to yield a
CC plasmid designated pC3FGAG. Termination codons were included to form
CC the plasmid pC3FGAGVQ. The Felv gag/pol expression cassette was
CC excised from this plasmid and designated pC3DOFGAGVQ. pC3DOFGAGVQ
CC was used in standard in vitro recombination assays with vCP83 and
CC vCP87 as rescue viruses. Recombinants containing the entire Felv-A
CC gag/pol sequences and the entire Felv-A env gene were designated
CC vCP97, while recombinants containing the same gag/pol sequences and
CC the entire Felv-A env lacking the immuno-suppressive region were
CC designated vCP93. The Felv-A gag can be inserted into a vaccinia
CC virus background by including a polylinker sequence flanking the
CC coding region and using these sequences to insert the gag region into
CC the A type inclusion body region (ATI) of the plasmid pSD541. See
CC also AA035501-864.

SO Sequence 3674 BP; 970 A; 1018 C; 883 G; 803 T; 0 other;

Alignment Scores:

Pred. No.:	373	Length:	3674
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.76%	Indels:	0
DB:	13	Gaps:	0

US-10-024-955-7 (1-213) x AA029114 (1-3674)

QY 196 LysGluMetThrLysValLeuAla 203

DB 2188 AAGGAGATGACATAAGTTCTGGCC 2211

RESULT 20
AA013061/C
ID AA013061 standard; DNA; 6027 BP.

XX
AC AA013061;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:124.

KW Enterococcus faecalis; contig; detection: Enterococcal infection;

KM vaccine; attenuation; computer readable medium; ds.

XX
OS Enterococcus faecalis.

PN WO9850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HDMA-) HDMA GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI; 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
infection.

PS Claim 1; Page 756-759; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AA012938 to AA013919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
infection.

SO Sequence 6027 BP; 1761 A; 1288 C; 999 G; 1970 T; 9 other;

Alignment Scores:

Pred. No.:	595	Length:	6027
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.76%	Indels:	0
DB:	20	Gaps:	0

US-10-024-955-7 (1-213) x AA013061 (1-6027)

QY 6 LeuIleAlaIaValAlaPheVal 13

DB 2072 TTAAATTGACGGGTGACATTGT 2049

RESULT 21

AB04928/C
ID AB04928 standard; cDNA; 7318 BP.

XX
AC AB04928;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 9266.

KW Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ss.

XX
OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB60825.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Claim 1; SEQ ID NO 9266; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC	Intestinal tissue of a cat following inoculation with lymphosarcoma
CC	cell-free supernatant from a cat with naturally-occurring feline
CC	Lymphosarcoma. Full-length provirus DNA was isolated by screening a
CC	library with an ITR-specific probe and subcloning into pUC18. Clone
CC	61E is replication competent and capable of inducing persistent
CC	viraemia. Host cells, e.g. feline cells, can be transformed with
CC	DNA derived from 61E and used to produce infectious virus useful in
CC	vaccines, in the generation of viraemia and in disease challenge
CC	systems.
SQ	Sequence 8440 BP; 2334 A; 2333 C; 1949 G; 1824 T; 0 other;
SO	
Alignment Scores:	
Pred. No.:	819
Score:	8.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	3.76%
Db:	17 Gaps: 0
US-10-024-955-7 (1-213) x AAT13265 (1-8440)	
OY	196 LysGIumEtThrLysValIeuAla 203
ID	
DB	2220 AAGGAGTACTAAAGTTCTGGC 2243
RESULT 23	
AKA81402	
ID	AKA81402 standard; DNA; 11623 BP.
XX	
AC	AKA81402;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36214.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer
KM	cyclostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
PD	
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 PS
 PS Disclosure; SEQ ID NO 36214; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM62170 to AAM91921. (I) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients' own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM62169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 11623 BP; 2524 A; 3007 C; 3158 G; 2934 T; 0 other;
 Alignment Scores: Length: 11623
 Pred. No.: 8.00 Matches: 8
 Score:

AC AAD32837;
XX
XX 01-JUL-2002 (first entry)
XX
DE Human FOXp3 1150G>A mutant gene.
XX
XX Human; detection; mutation; scurvy gene; FOXp3 gene; scurvy disease;
KW FOXp3 gene-related disease; X-linked disorder; polyendocrinopathy;
KM immune dysregulation; diagnosis; enteropathy; X-linked syndrome; mutant;
KM ds.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key location/Qualifiers
FH replace (1150, A)
FT /*tag= a
XX
XX WO200216656-A2.
XX
XX 28-FEB-2002.
XX
XX 20-AUG-2001; 2001WO-US41814.
XX
XX 21-AUG-2000; 2000US-226759P.
XX
XX (CELL-) CELTECH R & D INC.
XX
XX Brunkow ME;
XX
XX WPI: 2002-292072/33.
XX
XX
XX Detecting mutations of human orthologs of murine scurvy gene, FOXp3 for
PT diagnosing FOXp3 gene-related diseases in humans, by amplifying FOXp3
PT nucleic acid sequence using oligonucleotide primers and detecting
PT mutations -
XX
XX
XX Example 1; Page -: 40pp; English.
XX
XX The invention relates to methods and compositions for detecting a
CC mutation in a human orthologue of the murine scurvy gene, termed FOXp3.
CC The method is useful for detecting mutations of the FOXp3 gene and is
CC useful for diagnosis FOXp3 gene-related diseases in humans. Mutations
CC in the human scurvy/FOXp3 gene causing human X-linked disorders which
CC may or may not be similar to scurvy disease in mice, may be detected.
CC An e.g. of such a human disorder is immune dysregulation, enteropathy,
CC polyendocrinopathy or X-linked syndrome. The present sequence is human
CC FOXp3 1150G>A mutant gene.
CC Note: This sequence is not shown in the specification, however it is
CC constructed based on the human FOXp3 gene shown in SEQ.ID.NO:20
CC (AAD32823) of the sequence listing.
XX
XX
SQ Sequence 20000 BP; 4618 A; 6045 C; 4840 G; 4497 T; 0 other;
XX
XX
XX Alignment Scores:
Pred. No.: 1.85e+03 Length: 20000
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 24 Gaps: 0
XX
US-10-024-955-7 (1-213) x AAD32837 (1-20000)
XX
OY 73 MetAAGAsnIleGLuAlaArgly 80
DB 10940 ATGAGGAATATTGAGCGCAGGGG 10963
XX
XX
XX RESULT 27
XX AAH68530 standard; DNA; 349980 BP.
XX
XX AAH68530;
AC

XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 7065.
XX
XX
XX Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
KM
XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
XX
XX
XX 20-JUN-2001.
XX
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0158162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
XX
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for analysing
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
XX Disclosure; SEQ ID NO: 7065; 246pp + Sequence Listing; English.
XX
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX
XX
SQ Sequence 349980 BP; 86896 A; 98023 C; 80939 G; 84122 T; 0 other;
XX
XX
XX Alignment Scores:
Pred. No.: 2.77e+04 Length: 349980
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 22 Gaps: 0
XX
US-10-024-955-7 (1-213) x AAH68530 (1-349980)
XX
OY 8 AlaAlaValAlaPheValAlaVal 15
DB 265079 GCGGCGAGTTGCTTGTGCTGTC 265102
XX
XX
XX RESULT 28
XX AAT90312/C
ID AAT90312 standard; DNA; 30 BP.
XX
XX AAT90312;
AC
XX
XX 17-APR-1998 (first entry)
XX
XX 5' PCR primer 1.
DE

PF 15-OCT-1999; 99WO-US24205.
 XX
 PR 15-OCT-1998; 98US-0104435.
 XX
 PA (GENV) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Werberg D, Treacy M;
 DR WPI: 2000-317937/27.
 XX
 PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (seSTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX
 PS Claim 1; Page 302; 618pp; English.
 XX
 CC AAA11261 to AAA43419 represent specifically claimed secreted expressed
 CC sequence tags (seSTs), isolated from human, mouse, xenopus and rat
 CC tissue sources. The seSTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haemopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;
 CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The seSTs can be used for gene
 CC therapy and in vaccines. The seSTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the seSTs. Proteins encoded by the seSTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA4425 represent linker variants which are given
 CC in the exemplification of the present invention.
 XX
 SQ Sequence 211 BP; 43 A; 62 C; 45 G; 61 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 272 Length: 211
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.29% Indels: 0
 DB: 21 Gaps: 0
 US-10-024-955-7 (1-213) x AAA1893 (1-211)
 QY 13 ValAlaValSerAlaAspPro 19
 Db 127 GTTGAGTGAGTGACAGATCC 107
 RESULT 31
 AAQ73773/C
 ID AAQ73773 standard; cDNA; 222 BP.
 XX
 AC AAQ73773;
 XX
 DT 26-MAY-1995 (first entry)
 XX
 DE Hepatitis C virus type-specific sequence.
 XX
 KW HCV-4; HCV-5; HCV-6; NS4; NS5; non-structural protein; antigen;
 KW antibody; vaccine; diagnosis; therapy; typing; immunoassay; ss.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9425602-A.

XX
 PD 10-NOV-1994.
 XX
 PF 05-MAY-1994; 94WO-GB00957.
 XX
 PR 05-MAY-1993; 93GB-0009237.
 PR 07-JAN-1994; 94GB-0000263.
 XX
 PA (COMN-) COMMON SERVICES AGENCY.
 PA (MURE-) MUREX DIAGNOSTICS LTD.
 XX
 PI Pike IH, Simmonds P, Yap PL;
 DR WPI: 1994-358278/44.
 XX
 PT New polynucleotide(s) specific for hepatitis C virus types 4, 5
 PT and 6 - and related antigenic peptide(s) and antibodies, useful
 PT in vaccines, diagnosis, HCV typing and treatment
 XX
 PS Claim 11; Page 57; 70pp; English.
 XX
 CC Viral RNA from HCV-infected patients was subjected to PCR
 CC amplification using primers corresponding to highly conserved
 CC sequences of non-structural proteins NS5 and NS4. Products were
 CC cloned into pUC18 and expressed in Escherichia coli. NS5 cDNA
 CC sequences unique to HCV-4 and HCV-6, and NS4 sequences unique to
 CC HCV-4, HCV-5 and the newly discovered HCV-6 are given in AAQ73767-75.
 XX
 SQ Sequence 222 BP; 54 A; 60 C; 66 G; 42 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 286 Length: 222
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.29% Indels: 0
 DB: 15 Gaps: 0
 US-10-024-955-7 (1-213) x AAQ73773 (1-222)
 QY 6 LeuIleAlaValAlaPhe 12
 Db 197 CTATATGCGCGCGCTGCTTTT 177
 RESULT 32
 ABN68279
 ID ABN68279 standard; DNA; 261 BP.
 XX
 AC ABN68279;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Streptococcus polynucleotide seq ID NO 4471.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 OS Streptococcus pyogenes.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX

XX 21-APR-1998: 980S-082567P.
XX
XX (LALG/) LALGUDI R V.
PA (ITOL/) ITO L Y.
PA (SHER/) SHERMAN B K.
XX
PI Lalgudi RV, Ito LY, Sherman BK;
XX
XX WPI; 2002-163647/21.
XX
PT Novel purified corn tassel-derived polynucleotide useful for
PT determining altered gene expression, to recover regulatory elements and
PT to follow inheritance of desirable characteristics through hybrid
PT breeding programs -
XX
XX
PS Claim 1: SEQ ID 2975; 201pp; English.
XX
XX The present sequence describes a purified corn tassel-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76833. The cdps sequences
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (1)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (1) are also useful
CC in the evaluation, and alteration of desired characteristics associated
CC with growth and development, disease resistance, environmental
CC adaptability, quality and yield, and as molecular markers for studying
CC inheritance of multigene traits in a plant breeding program. (1) can be
CC used to produce a tassel-specific profile of gene transcription, a
CC transcript image, to clone regulatory elements for use in transformation
CC vectors, to express a polypeptide, to identify, isolate or extend
CC identical or related corn tassel nucleic acid sequences from DNA
CC libraries, in nucleic acid hybridisation or amplification technologies,
CC as query sequences to determine homology of known sequences, as probe
CC for use in Southern or Northern hybridisation, and to identify the
CC presence of and/or to determine the degree of similarity between two
CC (or more) nucleic acid sequences.
XX
SQ Sequence 289 BP; 75 A; 61 C; 62 G; 83 T; 8 other:

Alignment Scores:
Pred. No.: 367 Length: 289
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 24 Gaps: 0

US-10-024-955-7 (1-213) x ABL73601 (1-289)
OY 170 G1yG1yLeuSer11LeuAsp 176
Db 198 GCGGCGCTTTCATCTAGAC 178

RESULT 35
ID AAK78516/C
XX AAK78516 standard; DNA: 290 BP.
AC AAK78516:
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33328.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX
XX MO200157182-A2.
PN
XX
PD 09-AUG-2001.

XX
XX 17-JAN-2001: 2001WO-US01354.
XX
XX 31-JAN-2000: 2000US-0179065.
PR 04-FEB-2000: 2000US-0180628.
PR 24-FEB-2000: 2000US-0184664.
PR 02-MAR-2000: 2000US-0186350.
PR 16-MAR-2000: 2000US-0189874.
PR 17-MAR-2000: 2000US-0190076.
PR 18-APR-2000: 2000US-0198123.
PR 19-MAY-2000: 2000US-0205515.
PR 07-JUN-2000: 2000US-0209467.
PR 28-JUN-2000: 2000US-0214886.
PR 30-JUN-2000: 2000US-0215135.
PR 07-JUL-2000: 2000US-0216647.
PR 07-JUL-2000: 2000US-0216880.
PR 11-JUL-2000: 2000US-0217487.
PR 11-JUL-2000: 2000US-0217496.
PR 14-JUL-2000: 2000US-0218290.
PR 26-JUL-2000: 2000US-0220963.
PR 26-JUL-2000: 2000US-0220964.
PR 14-AUG-2000: 2000US-0224518.
PR 14-AUG-2000: 2000US-0224519.
PR 14-AUG-2000: 2000US-0225213.
PR 14-AUG-2000: 2000US-0225214.
PR 14-AUG-2000: 2000US-0225266.
PR 14-AUG-2000: 2000US-0225267.
PR 14-AUG-2000: 2000US-0225268.
PR 14-AUG-2000: 2000US-0225270.
PR 14-AUG-2000: 2000US-0225447.
PR 14-AUG-2000: 2000US-0225757.
PR 14-AUG-2000: 2000US-0225758.
PR 14-AUG-2000: 2000US-0225759.
PR 18-AUG-2000: 2000US-0226279.
PR 22-AUG-2000: 2000US-0226681.
PR 22-AUG-2000: 2000US-0226688.
PR 22-AUG-2000: 2000US-0227182.
PR 23-AUG-2000: 2000US-0227009.
PR 30-AUG-2000: 2000US-0228924.
PR 01-SEP-2000: 2000US-0229287.
PR 01-SEP-2000: 2000US-0229343.
PR 01-SEP-2000: 2000US-0229344.
PR 01-SEP-2000: 2000US-0229345.
PR 05-SEP-2000: 2000US-0229509.
PR 05-SEP-2000: 2000US-0229513.
PR 06-SEP-2000: 2000US-0230437.
PR 06-SEP-2000: 2000US-0230438.
PR 08-SEP-2000: 2000US-0231242.
PR 08-SEP-2000: 2000US-0231243.
PR 08-SEP-2000: 2000US-0231244.
PR 08-SEP-2000: 2000US-0231413.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0232080.
PR 08-SEP-2000: 2000US-0232081.
PR 12-SEP-2000: 2000US-0231968.
PR 14-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232398.
PR 14-SEP-2000: 2000US-0232399.
PR 14-SEP-2000: 2000US-0232400.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0233063.
PR 14-SEP-2000: 2000US-0233064.
PR 14-SEP-2000: 2000US-0233065.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234997.
PR 25-SEP-2000: 2000US-0234998.
PR 26-SEP-2000: 2000US-0234984.
PR 27-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235836.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236367.
PR 29-SEP-2000: 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI, 2001-483426/52.
 XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 XX
 PS Disclosure: SEQ ID NO 33328; 3071bp + Sequence Listing: English.
 XX
 CC AKK4951 to AKK64702 encode the human immune/hematopoietic antigen (1)
 CC amino acid sequences given in AAM82170 to AAM91921. (1) have cytostratic
 CC activity, and can be used in gene therapy and vaccine production. (1)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (1) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (1) by expressing inactive proteins or to
 CC supplement the patient's own production of (1). Additionally, (1)
 CC polynucleotides may be used to produce the secreted (1), by inserting the
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (1) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AKK64703
 CC to AKK7694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AKK54942 to AKK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 290 BP; 57 A; 72 C; 76 G; 85 T; 0 other;
 Alignment Scores:
 Pred. No.: 368 Length: 290
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.29% Indels: 0
 DB: 22 Gaps: 0
 US-10-024-955-7 (1-213) x AAK78516 (1-290)
 OY 142 LeuSerLeuGluIleSerasp 148
 DB 45 CTGCTCTGGAATCATGTCAC 25
 RESULT 36
 AAV86640/C
 ID AAV86640 standard; cDNA; 294 BP.
 AC AAV86640;
 XX
 DT 27-APR-1999 (first entry)
 XX
 DE EST clone AX128.
 XX
 KW Expressed sequence tag; secreted protein; hematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO9845435-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06954.
 XX
 PR 10-APR-1997; 97US-0835913.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 XX
 XX Racie LA, Spaulding V, Treacy M;
 DR WPI, 1999-070076/06.
 XX
 PT New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 301; 633pp; English.
 XX
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.
 CC
 SQ Sequence 294 BP; 97 A; 66 C; 66 G; 65 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 373 Length: 294
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.29% Indels: 0
 DB: 20 Gaps: 0
 US-10-024-955-7 (1-213) x AAV86640 (1-294)
 QY 181 ValLeuSerAspValLeuThr 187
 Db 243 GTGTATCTATCATGCTTGACA 223
 RESULT 37
 ABK74171
 ID ABK74171 standard; DNA: 306 BP.
 XX
 AC ABK74171:
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Bacillus licheniformis genomic sequence tag (GST) #1462.
 XX
 KW Differential gene expression; genomic sequenced tag; GST;
 KW altered culture condition; environmental stress;
 KW physiological provocation; ds.
 XX
 OS Bacillus licheniformis.
 XX
 PN WO200229113-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 05-OCT-2001; 2001WO-US31437.
 XX
 PR 06-OCT-2000; 2000US-0680598.
 PR 27-MAR-2001; 2001US-279526P.
 XX
 PA (NOVO) NOVOZYMES BIOTECH INC.
 PA (NOVO) NOVOZYMES AS.
 XX
 PI Berka R, Clausen IG;
 XX
 DR WPI; 2002-416684/44.
 XX
 PT Monitoring differential expression of several genes in first Bacillus
 PT cell relative to expression of same genes in one or more second
 PT Bacillus cells, by using substrate containing Bacillus genomic
 PT sequenced tag array -
 XX

PS Claim 4; SEQ ID NO 1462; 200pp; English.
 XX
 CC The invention describes a method of monitoring differential expression of
 CC genes in a first Bacillus cell relative to expression of the genes in
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
 CC isolated from Bacillus cells to a substrate containing array of Bacillus
 CC genomic sequenced tags (GST), examining the array, and determining
 CC relative gene expression by an observed hybridisation reporter signal of
 CC a spot in the array. The method is useful for measuring the expression of
 CC genes in a first Bacillus cell relative to expression of the same genes
 CC in one or more second Bacillus cells. The method is useful for monitoring
 CC global expression of several genes from a Bacillus cell, discovering new
 CC genes, identifying possible functions of unknown open reading frames and
 CC monitoring gene copy number variation and stability. Monitoring changes
 CC in expression of genes may be used to provide a representation of the way
 CC in which Bacillus cells adapt to changes in culture conditions,
 CC environmental stress or other physiological provocation. Extensive
 CC follow-up characterisation is unnecessary, when one spot on an array
 CC equals one gene or one open reading frame, since sequence information is
 CC available. This sequence represents a genomic sequence tag (GST) used in
 CC the method of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 306 BP; 96 A; 62 C; 83 G; 65 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 387 Length: 306
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.29% Indels: 0
 DB: 24 Gaps: 0
 US-10-024-955-7 (1-213) x ABK74171 (1-306)
 QY 197 GluMetThrIysValLeuAla 203
 Db 72 GAATATGACAAAGTGTGCG 92
 RESULT 38
 AAV77349
 ID AAV77349 standard; DNA: 334 BP.
 XX
 AC AAV77349:
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE Staphylococcus aureus contig SEQ ID #3038.
 XX
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 XX
 OS Staphylococcus aureus.
 XX
 PN EP786519-A2.
 XX
 PD 30-JUL-1997.
 XX
 PF 07-JAN-1997; 97EP-0100117.
 XX
 PR 05-JAN-1996; 96US-0009861.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 XX
 DR WPI; 1997-374922/35.
 XX

XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus* stored on computer readable medium and used in the production of anti-*S. aureus* vaccines

PS Claim 1; Page 2418; 3271pp; English.

XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the *S. aureus* DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against *S. aureus* infection. The polypeptides can also be used in a kit for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the *S. aureus* DNA sequences contained on the computer readable medium.

XX Sequence 334 BP; 87 A; 57 C; 70 G; 117 T; 3 other:

SO Alignment Scores:

Pred. No.:	421	Length:	334
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	18	Gaps:	0

US-10-024-955-7 (1-213) x AAV77349 (1-334)

OY 141 AlaleuSerleuGluIleSer 147
|||||

DB 177 GCGCTGTCCTGAGAAATATCT 197

RESULT 39
AAK95608
ID AAK95608 standard; DNA; 401 BP.

AC AAK95608;

XX 17-DEC-2001 (first entry)

DE Human neutregulin gene single nucleotide polymorphism SNP8NRG383932.

XX Human; neutregulin-1 associated gene 1; NRG1AG1; Schizophrenia gene; gene therapy; single nucleotide polymorphism; SNP; ds.

OS Homo sapiens.

XX WO200164876-A2.

PN 07-SEP-2001.

PD 28-FEB-2001; 2001WO-US06376.

PF 28-FEB-2000; 2000US-0515715.

PR (DECO-) DECODE GENETICS EHF.

PA Stefansson H, Steinthorsdottir V, Gulcher JR.

PI WPI: 2001-550179/61.

XX Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for preventing diagnosing and treating schizophrenia

XX Disclosure: Page 561; 750pp; English.

PS This sequence represents a single nucleotide polymorphism (SNP) of the human neutregulin-1 associated gene 1 (NRG1AG1) of the invention. The NRG1AG1 gene is also referred to as the human schizophrenia gene. The invention also relates to fragments or variants of the gene and the NRG1AG1 polypeptides they encode. The NRG1AG1 nucleic acids and polypeptides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate NRG1AG1 expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of NRG1AG1 by expressing inactive proteins or to supplement the patients own production of NRG1AG1. Additionally, the gene may be used to produce NRG1AG1 polypeptides, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The gene may also be used as DNA probes and primers in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The NRG1AG1 polypeptides may also be used as antigens in the production of antibodies against NRG1AG1 and in assays to identify modulators of NRG1AG1 expression and activity. Anti-NRG1AG1 antibodies and antagonists may also be used to down regulate expression and activity. Anti-NRG1AG1 antibodies may also be used as diagnostic agents for detecting the presence of NRG1AG1 polypeptides in samples. NRG1AG1 is associated with schizophrenia which may be prevented, diagnosed and/or treated by the above methods.

XX Sequence 401 BP; 99 A; 71 C; 95 G; 134 T; 2 other:

SO Alignment Scores:

Pred. No.:	500	Length:	401
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	22	Gaps:	0

US-10-024-955-7 (1-213) x AAK95608 (1-401)

OY 102 HisLeuLeuIleGlyValHis 108
|||||

DB 357 CATCTGTGATGCGGCTTCAT 377

RESULT 40
AAK97101
ID AAK97101 standard; DNA; 401 BP.

AC AAK97101;

XX 17-DEC-2001 (first entry)

DE Human neutregulin gene single nucleotide polymorphism SNP8NRG383932.

XX Human; neutregulin 1 gene; schizophrenia; gene therapy; SNP; single nucleotide polymorphism; ds.

OS Homo sapiens.

XX WO200164877-A2.

PN 07-SEP-2001.

PD 28-FEB-2001; 2001WO-US06377.

PF 28-FEB-2000; 2000US-0515716.

PR (DECO-) DECODE GENETICS EHF.

PA Stefansson H, Steinthorsdottir V, Gulcher JR.

PI WPI: 2001-514841/56.

XX

PT Neuregulin 1 nucleic acids and proteins useful for diagnosing
PT preventing and treating schizophrenia -
XX
PS Disclosure: Page 146; 756bp; English.
XX
CC This sequence represents a single nucleotide polymorphism (SNP)
CC from the human neuregulin 1 gene of the invention.
CC The invention also relates to fragments or variants of the neuregulin 1
CC gene. The gene and its proteins may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate neuregulin 1
CC expression, such as schizophrenia. For example they may be used to treat
CC disorders associated with decreased neuregulin 1 expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC neuregulin 1 by expressing inactive proteins or to supplement the
CC patients own production of polypeptides. Additionally, the gene may be
CC used to produce the neuregulin 1 protein, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. The gene
CC and its complementary sequences may also be used as DNA probes in
CC diagnostic assays to detect and quantitate the presence of similar
CC nucleic acids in samples, and therefore which patients may be in need of
CC restorative therapy. The protein may also be used as antigens in the
CC production of antibodies against neuregulin 1 and in assays to identify
CC modulators of neuregulin 1 expression and activity. The antibodies and
CC antagonists may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of neuregulin 1 in samples.
XX
SQ Sequence 401 BP; 99 A; 71 C; 95 G; 134 T; 2 other;
XX
Alignment Scores:
Pred. No.: 500 Length: 401
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 22 Gaps: 0
US-10-024-955-7 (1-213) x AAK97101 (1-401)
QY 102 HisleuLeuIleGlyValHis 108
DB 357 CATCTGTGATTGGGTTTCAT 377
XX
RESULT 41
AAI90301
ID AAI90301 standard; cDNA; 408 BP.
XX
AC AAI90301:
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 10361.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX

DR WPI; 2001-514838/56.
DR P-PSDB; AAO10370.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 10361; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 408 BP; 145 A; 84 C; 107 G; 71 T; 1 other;
XX
Alignment Scores:
Pred. No.: 508 Length: 408
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 22 Gaps: 0
US-10-024-955-7 (1-213) x AAI90301 (1-408)
QY 203 AlaProAlaPheIysArgGlu 209
DB 124 GCCCCACGCTTTTAAAGGAA 144
XX
RESULT 42
AAK62472
ID AAK62472 standard; cDNA; 419 BP.
XX
AC AAK62472:
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:7532.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PR 04-FEB-2000; 2000US-0180628.
XX
PR 24-FEB-2000; 2000US-0184654.
XX
PR 02-MAR-2000; 2000US-0186350.
XX
PR 16-MAR-2000; 2000US-0189874.
XX
PR 17-MAR-2000; 2000US-0190076.
XX
PR 18-APR-2000; 2000US-0198123.
XX
PR 19-MAY-2000; 2000US-0205515.
XX
PR 07-JUN-2000; 2000US-0209467.
XX
PR 28-JUN-2000; 2000US-0214886.
XX
PR 30-JUN-2000; 2000US-0215135.
XX
PR 07-JUL-2000; 2000US-0216647.
XX

PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235837.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0235678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.
P-PSDB; AAM89691.
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -
Claim 1; SEQ ID NO 7532; 3071pp + Sequence Listing; English.
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
activity, and can be used in gene therapy and vaccine production. (II)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to

CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 419 BP; 94 A; 105 C; 122 G; 90 T; 8 other;

Alignment Scores:
Pred. No.: 521 Length: 419
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 22 Gaps: 0

US-10-024-955-7 (1-213) x AAK62472 (1-419)

QY 187 ThrAlarIlePheGlnAspThr 193

DB 36 ACAGCCATCTTTCAAGACACC 56

RESULT 43

ID ABL15457/C standard; cDNA; 446 BP.

XX ABL15457;

DT 13-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 15448.

XX Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX MO200160860-A2.

XX 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 23-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 2591; 11750pp; English.

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 446 BP; 178 A; 98 C; 69 G; 101 T; 0 other;

Alignment Scores:
Pred. No.: 553 Length: 446
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 23 Gaps: 0

US-10-024-955-7 (1-213) x ABL15457 (1-446)

QY 171 GlyLeuSerIleLeuAspPro 177

DB 387 GGGTTAAGTATTTTACACCT 367

RESULT 44

ID ABL83546 standard; cDNA; 471 BP.

XX ABL83546;

DT 17-MAY-2002 (first entry)

XX Human ovarian cancer related cDNA clone SEQ ID NO:6524.

XX Human: ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

XX MO200192581-A2.

XX 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US17756.

PR 26-MAY-2000; 2000US-207484P.

XX (COR-) CORIXA CORP.

PI Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising

PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding

PT polypeptide, antibody specific to polypeptide or T cell expressing

XX Claim 1; SEQ ID 6524; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers

CC and immunostimulants; and a polypeptide (II) of a ovarian tumour

CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence

CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to

CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell

CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is

CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
SQ Sequence 471 BP; 128 A; 111 C; 84 G; 148 T; 0 other;
Alignment Scores:
Pred. No.: 582 Length: 471
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 24 Gaps: 0
US-10-024-955-7 (1-213) x ABL83546 (1-471)
QY 177 ProIlePheGlyValIleuser 183
DB 65 CCCATCTTGGAGTCTCA 85
RESULT 45
ABL13765/C
ID ABL13765 standard; cDNA; 489 BP.
XX
XX ABL13765;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 35777.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KM
XX
XX pharmaceutical; gene; ss.
OS
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PE
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX P-PSDB; ABB69662.
DR
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX
XX Claim 1: SEQ ID NO 35777; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB12072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 489 BP; 141 A; 120 C; 138 G; 90 T; 0 other;
Alignment Scores:
Pred. No.: 603 Length: 489
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 23 Gaps: 0
US-10-024-955-7 (1-213) x ABL13765 (1-489)
QY 8 AlaAlaValAlaPheValAla 14
DB 408 GCTGCAGTTGGCTTTGTCCCA 388
RESULT 46
AAC94251/C
ID AAC94251 standard; cDNA; 500 BP.
XX
XX AAC94251;
AC
XX
XX 19-FEB-2001 (first entry)
DT
XX
XX Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:746.
DE
XX
XX Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KM
XX
XX vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
XX Ctenocephalides felis.
OS
XX
XX WO200061621-A2.
PN
XX
XX 19-OCT-2000.
PD
XX
XX 07-APR-2000; 2000WO-US09437.
PE
XX
XX 09-APR-1999; 99US-0128704.
PR
XX
XX (HESK-) HESKA CORP.
PA
XX
XX Brandt KS, Gaines PJ, Stinchcomb DF, Wisniewski N;
PI
XX
XX WPI; 2000-656323/63.
DR
XX
XX Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations -
PT
XX
XX Claim 26: Page 489; 964pp; English.
PS
XX
XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The

CC anti-HMT/HMC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HMC cDNA of the invention.

XX Sequence 500 BP; 169 A; 86 C; 67 G; 178 T; 0 other;

Alignment Scores:

Pred. No.:	616	Length:	500
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	21	Gaps:	0

US-10-024-955-7 (1-213) x AAC94251 (1-500)

QY 136 GlnAspPheValAlaLeu 142

DB 165 CAGGATTTGTGTGGCCTTG 145

RESULT 47

AKI1020/C
ID AKI1020 standard; DNA; 501 BP.

XX AAK1020;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 11011.

XX Human: brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer; ss.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR.

PI WPI; 2001-483446/52.

DR WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

PT brains -

PS Example 4; SEQ ID NO: 11011; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is one of the probes of the

CC invention.

XX Sequence 501 BP; 212 A; 81 C; 101 G; 107 T; 0 other;

Alignment Scores:

Pred. No.:	617	Length:	501
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	22	Gaps:	0

US-10-024-955-7 (1-213) x AKI1020 (1-501)

QY 186 LeuThrAlaIlePheGlnAsp 192

DB 268 CTACTGCTATTTCACAGAT 248

RESULT 48

AAI17708/C
ID AAI17708 standard; DNA; 501 BP.

XX AAI17708;

DT 12-OCT-2001 (first entry)

DE Probe #7641 for gene expression analysis in human cervical cell sample.

XX Probe: human; microarray; gene expression; cervical epithelial cell;

KW cervical cancer; ss.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR.

PI WPI; 2001-488901/53.

DR WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 7641; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes

CC (SENPs). The present sequence is one such probe. The SENPs are derived

CC from human HeLa cells. The SENPs can be used to produce a single exon

CC microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene

CC expression, the probes are therefore useful in grading and/or staging

CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 501 BP; 212 A; 81 C; 101 G; 107 T; 0 other;

XX Alignment Scores:

Pred. No.:	617	Length:	501
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0


```

DB:                22                Gaps:                0
US-10-024-955-7 (1-213) x AAI17708 (1-501)
OY      186 LeuThra1a1lePheglnasp 192
        |||||||
DB      268 CTTACTGCTATTTCACAGAT 248

RESULT 49
AAI42644/c
ID      AAI42644 standard; DNA; 501 BP.
XX
XX      AAI42644;
AC
XX      17-OCT-2001 (first entry)
DT
XX      Probe #11330 used to measure gene expression in human placenta sample.
DE
XX      Probe; microarray; human; placenta; antenatal diagnosis;
KW      genetic disorder; ss.
XX
XX      Homo sapiens.
OS
XX      WO200157272-A2.
PN
XX      09-AUG-2001.
PD
XX      30-JAN-2001; 2001WO-US000663.
PF
XX      04-FEB-2000; 2000US-0180312.
PR
XX      26-MAY-2000; 2000US-0207456.
PR
XX      30-JUN-2000; 2000US-0608408.
PR
XX      03-AUG-2000; 2000US-0632366.
PR
XX      21-SEP-2000; 2000US-0234687.
PR
XX      27-SEP-2000; 2000US-0236359.
PR
XX      04-OCT-2000; 2000GB-0024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX      WPI; 2001-488897/53.
DR
XX
XX      Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human placenta -
CC
XX      Claim 25; SEQ ID NO 11330; 654pp; English.
XX
XX      The present invention relates to single exon nucleic acid probes (SENP).
CC      The present sequence is one such probe. The probes are useful for
CC      producing a microarray for predicting, measuring and displaying gene
CC      expression in samples derived from human placenta. The probes are useful
CC      for antenatal diagnosis of human genetic disorders.
XX
XX      Sequence 501 BP; 212 A; 81 C; 101 G; 107 T; 0 other;
SQ
XX
XX      Alignment Scores:
XX      Pred. No.:          617          Length:          501
XX      Score:              7.00          Matches:          7
XX      Percent Similarity: 100.00%      Conservative:    0
XX      Best Local Similarity: 100.00%      Mismatches:     0
XX      Query Match:         3.29%        Indels:          0
XX      DB:                  22          Gaps:            0
US-10-024-955-7 (1-213) x AAI42644 (1-501)
OY      186 LeuThra1a1lePheglnasp 192
        |||||||
DB      268 CTTACTGCTATTTCACAGAT 248

RESULT 50
AAC01139
ID      AAC01139 standard; cDNA; 512 BP.

```

```

XX      AAC01139;
AC
XX      06-OCT-2000 (first entry)
DT
XX      Human secreted protein 5' EST, SEQ ID NO: 1137.
DE
XX      Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW      gene therapy; chromosome mapping; ss.
XX
XX      Homo sapiens.
OS
XX      EP1033401-A2.
PN
XX      06-SEP-2000.
PD
XX      21-FEB-2000; 2000EP-0200610.
PF
XX      26-FEB-1999; 99US-0122487.
PR
XX      (GEST ) GENSET.
PA
XX      Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX      WPI; 2000-500381/45.
DR
XX      P-PSDB: AAG01133.
DR
XX
XX      New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT      obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT      diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS
XX      Claim 1; SEQ ID 1137; 71pp + CD-ROM; English.
XX
XX      The present sequence is one of a large number of 5' ESTs derived from
CC      cDNAs encoding secreted proteins. An ORF has been identified within the
CC      sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC      derived from 30 different tissues. EST sequences usually correspond
CC      mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC      often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC      well suited for isolating cDNA sequences derived from the 5' ends of
CC      mRNAs and even in those cases where longer cDNA sequences have been
CC      obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC      mRNAs with intact 5' ends and can therefore be used to obtain full length
CC      cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC      gene therapy and chromosome mapping procedures. They are used to obtain
CC      upstream regulatory sequences and to design expression and secretion
CC      vectors.
XX
XX      Sequence 512 BP; 113 A; 146 C; 171 G; 77 T; 5 other;
SQ
XX
XX      Alignment Scores:
XX      Pred. No.:          630          Length:          512
XX      Score:              7.00          Matches:          7
XX      Percent Similarity: 100.00%      Conservative:    0
XX      Best Local Similarity: 100.00%      Mismatches:     0
XX      Query Match:         3.29%        Indels:          0
XX      DB:                  21          Gaps:            0
US-10-024-955-7 (1-213) x AAC01139 (1-512)
OY      13 ValAlaValSerAlaAspPro 19
        |||||||
DB      287 GTGGCCGTTTCGACACCC 307

```

Search completed: February 20, 2003, 18:24:12
Job time : 367 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 18:34:51 ; Search time 78 Seconds
(without alignments)
837.463 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 213
Sequence: 1 MMKFLIAAVAFVAVSADPL.....VRKEMTKVLAPAFKRELEKN 213

Scoring table:

	OLIGO
Xgapop 60.0 ,	Xgapext 60.0
Ygapop 60.0 ,	Ygapext 60.0
Fgapop 6.0 ,	Fgapext 7.0
Delop 6.0 ,	Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878587

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPIO.spool/US10024955/runat_14022003_130736_24281/app_query.fasta.1.391
-DB=Issued Patents_NA -QEMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=50 -MODE=LOCAL
-OUTMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10024955 @CGN 1.1.61 @runat_14022003_130736_24281 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAR -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEX=7

Database :

- 1: Issued Patents_NA:*
- 2: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213	100.0	761	4	US-08-553-336A-6
2	25	11.7	812	4	US-08-462-778-1
3	25	11.7	812	4	US-08-553-336A-1
4	8	3.8	24	4	US-08-553-336A-12
5	8	3.8	3022	4	US-09-232-278A-8
6	8	3.8	3674	1	US-08-105-483-324
7	8	3.8	3674	1	US-08-709-209-324
8	8	3.8	3674	1	US-08-458-101-324
9	8	3.8	24	3	US-08-462-778-5
10	7	3.3	24	4	US-08-553-336A-5
11	7	3.3	30	2	US-08-620-969-1
12	7	3.3	594	1	US-08-404-732A-3

c 13	7	3.3	912	4	US-09-134-001C-1944	Sequence 1944, App
c 14	7	3.3	963	1	US-08-361-467B-1	Sequence 1, Appli
c 15	7	3.3	963	1	US-08-484-332C-1	Sequence 1, Appli
c 16	7	3.3	1644	3	US-08-948-564-5	Sequence 5, Appli
c 17	7	3.3	1757	2	US-08-453-848-14	Sequence 14, Appli
c 18	7	3.3	1757	2	US-09-169-027-14	Sequence 14, Appli
c 19	7	3.3	1793	2	US-08-453-848-6	Sequence 6, Appli
c 20	7	3.3	1793	4	US-09-169-027-6	Sequence 6, Appli
c 21	7	3.3	1853	1	US-08-404-732A-6	Sequence 6, Appli
c 22	7	3.3	2120	1	US-08-404-732A-4	Sequence 4, Appli
c 23	7	3.3	2224	1	US-08-404-732A-8	Sequence 8, Appli
c 24	7	3.3	2273	2	US-08-788-539A-1	Sequence 1, Appli
c 25	7	3.3	2799	2	US-08-874-186-42	Sequence 42, Appli
c 26	7	3.3	2959	4	US-09-449-285A-3	Sequence 3, Appli
c 27	7	3.3	4558	1	US-08-309-512-2	Sequence 2, Appli
c 28	7	3.3	4558	5	PCT-US92-08756A-2	Sequence 2, Appli
c 29	7	3.3	5517	4	US-09-120-663-1	Sequence 1, Appli
c 30	7	3.3	7886	4	US-09-453-702B-102	Sequence 102, App
c 31	7	3.3	8195	4	US-08-961-527-94	Sequence 94, Appli
c 32	7	3.3	48974	4	US-08-920-422-17	Sequence 17, Appli
c 33	7	3.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 34	7	3.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c 35	6	2.8	18	3	US-08-462-778-3	Sequence 3, Appli
c 36	6	2.8	18	4	US-08-553-336A-3	Sequence 3, Appli
c 37	6	2.8	18	4	US-08-553-336A-15	Sequence 15, Appli
c 38	6	2.8	20	4	US-09-297-535-4	Sequence 4, Appli
c 39	6	2.8	21	1	US-08-479-852-43	Sequence 43, Appli
c 40	6	2.8	21	1	US-08-479-852-123	Sequence 123, App
c 41	6	2.8	21	2	US-08-462-646-43	Sequence 43, Appli
c 42	6	2.8	21	2	US-08-462-646-123	Sequence 123, App
c 43	6	2.8	21	4	US-09-013-406-43	Sequence 43, Appli
c 44	6	2.8	21	4	US-09-013-406-123	Sequence 123, App
c 45	6	2.8	22	1	US-08-379-078-654	Sequence 654, App

ALIGNMENTS

RESULT 1
US-08-553-336A-6
Sequence 6, Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032Cp2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 761 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 43..681
; US-08-553-336A-6

Alignment Scores:
Pred. No.: 6.99e-220 Length: 761
Score: 213.00 Matches: 213
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x US-08-553-336A-6 (1-761)

QY 1 MetMetLysPheLeuLeuLeaAlaValAlaPheValAlaValSerAlaAspProIle 20
DB 43 ATGATGAAATTTTGTGTGATGCTGCCGTGCAATTTGTCGCCGTTCGCTGATCCAAAT 102
QY 21 HisTyrAspLysIleThrGluGluLeuLysAlaIleAspAspAlaIleAlaIle 40
DB 103 CACATGATATAAATCACCGAAGAAATCAACAAGCTATTGATGATGCCATTGCGCTATT 162
QY 41 GluGlnSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGluArg 60
DB 163 GAACCAATCCGAACCAATAGATCCAAATGAAGTACTGATCATGCCGATTAATTCGAACGT 222
QY 61 HisValGlyIleValAspPheLysGlyGluLeuAlaMetLysGlnIleGluAlaArgGly 80
DB 223 CATGTTGCTATTGTGGATTTCAAAAGGTGAATTAAGCCATGCCAAACATTGAAGCTCGAGA 282
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys 100
DB 283 TTGAAACAAATGAAAGCTCAAGGTGATGCTAATGTCAAAGTGAAGGGATTTGTTAAA 342
QY 101 AlaHisIleLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
DB 343 GCTCATTTGTTGATCGGTGTTCAAGATGATGCTCTCGATGGAATGATTTAGCATATC 402
QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAspIleGlnAspPheValVal 140
DB 403 AAATTGGGTATCTTCAATCCAACCATCATCATCTTCGGAATTCAGATTGTTGTT 462
QY 141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArg 160
DB 463 GCCTTGCCCTTGAAATTTTGTGATGAAGGTAACTAATGACATCTTTTGAAGTACGA 522
QY 161 GluPheAlaAsnValValAsnHisIleGlyLysSerIleLeuAspProIlePheGly 180
DB 523 CAATTCCTGATGTTGCAACCAATATGTTGCTTCAATCTTGATCCAAATTTTGGC 582
QY 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys 200
DB 583 GTTTTATCTGATGATTGACCGCTATTTCACAGACACCGTACCTAAGAAATGACCAAA 642
QY 201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
DB 643 GTATTGGCACACGATTTAAACGTAATTGGAATAAAT 681

RESULT 2
US-08-462-778-1
; Sequence 1, Application US/08462778
; Patent No. 6077517
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kaw-Yan
; TITLE OF INVENTION: Allergenic Protein and Peptides From

```

```

; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,778
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/031,141
; FILING DATE: 12 March 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 812 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 68..712
; US-08-462-778-1

Alignment Scores:
Pred. No.: 6.74e-18 Length: 812
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.74% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x US-08-462-778-1 (1-812)

QY 155 ThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIle 174
DB 530 ACATCGTTCCAGAGTACTCAATTTGGCAATGTTGTCATATCATATGTTGTCATTCAAT 589
QY 175 LeuAspProIlePhe 179
DB 590 TTGGATCAATTTTC 604

RESULT 3
US-08-553-336A-1
; Sequence 1, Application US/08553336A
; Patent No. 6413738
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA

```

```

;
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMT-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 812 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 68..712
; NAME/KEY: mat_peptide
; LOCATION: 119..712
;
; US-08-553-336A-1
;
; Alignment Scores:
; Pred. No.: 6.74e-18 Length: 812
; Score: 25.00 Matches: 25
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 11.74% Indels: 0
; DB: 4 Gaps: 0
;
; US-10-024-955-7 (1-213) x US-08-553-336A-1 (1-812)
;
; QY 155 ThrSerPheGluValArgGlnPheAlaAsnValAlaHisIleGlyGlyLeuSerIle 174
; DB 530 ACATGCTTCGAGTACGTCAATTGGCCAAATTGTCATCATATTTGGTCTTTCAAATT 589
; QY 175 LeuAspProIlePhe 179
; DB 590 TTGCATCCAAATTTTC 604
;
; RESULT 4
; US-08-553-336A-12
; Sequence 12, Application US/08553336A
; Patent No. 6413738
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

;
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMT-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
;
; US-08-553-336A-12
;
; Alignment Scores:
; Pred. No.: 0.437 Length: 24
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 3.76% Indels: 0
; DB: 4 Gaps: 0
;
; US-10-024-955-7 (1-213) x US-08-553-336A-12 (1-24)
;
; QY 173 SerIleLeuAspProIlePheGly 180
; DB 1 TCATCTTGATCCAAATTTTGGC 24
;
; RESULT 5
; US-09-232-278A-8
; Sequence 8, Application US/09232278A
; Patent No. 6348196
; GENERAL INFORMATION:
; APPLICANT: AUDONNET et al.
; TITLE OF INVENTION: FELINE POLYNUCLEOTIDE VACCINE FORMULA
; FILE REFERENCE: 454313-2220
; CURRENT APPLICATION NUMBER: US/09/232,278A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3022
; TYPE: DNA
; ORGANISM: Feline leukemia virus
; US-09-232-278A-8
;
; Alignment Scores:
; Pred. No.: 42.8 Length: 3022
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 3.76% Indels: 0
; DB: 4 Gaps: 0
;
; US-10-024-955-7 (1-213) x US-09-232-278A-8 (1-3022)
;
; QY 196 LysGluMetThrLysValLeuAla 203
; DB 1537 AAGGAGATGACATTAAGTCTGGCC 1560
;
; RESULT 6
; US-08-105-483-324
; Sequence 324, Application US/08105483

```

```
; Patent No. 5494807
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE
; TITLE OF INVENTION: STRAIN
; NUMBER OF SEQUENCES: 462
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; ADDRESSEE: c/o William S. Frommer
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/105,483
; FILING DATE: 12-AUG-1993
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/847,951
; FILING DATE: 06-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 324:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-105-483-324
;
; Alignment Scores:
; Pred. No.: 51.5 Length: 3674
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 3.76% Indels: 0
; DB: 1 Gaps: 0
;
; US-10-024-955-7 (1-213) x US-08-105-483-324 (1-3674)
;
; QY 196 Lysglumethrlysvallena1a 203
; Db 2188 AAGGAGTGAAGTTCGCCC 2211
;
; RESULT 7
; US-08-709-209-324
; Sequence 324, Application US/08709209
; Patent No. 5762938
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE
; TITLE OF INVENTION: STRAIN
; NUMBER OF SEQUENCES: 462
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; ADDRESSEE: c/o William S. Frommer
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,209
; FILING DATE: 21-AUG-1996
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/105,483
; FILING DATE: 12-AUG-1993
; APPLICATION NUMBER: US 07/847,951
; FILING DATE: 06-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 324:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-709-209-324
;
; Alignment Scores:
; Pred. No.: 51.5 Length: 3674
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 3.76% Indels: 0
; DB: 1 Gaps: 0
;
; US-10-024-955-7 (1-213) x US-08-709-209-324 (1-3674)
;
; QY 196 Lysglumethrlysvallena1a 203
; Db 2188 AAGGAGTGAAGTTCGCCC 2211
;
; RESULT 8
; US-08-458-101-324
; Sequence 324, Application US/08458101
; Patent No. 5765599
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Perkus, Marion E.
; APPLICANT: Taylor, Jill
; APPLICANT: Tartaglia, James
; APPLICANT: No. 5765599ton, Elizabeth K.
; APPLICANT: Riviere, Michel
; APPLICANT: de Taisne, Charles
; APPLICANT: Limbach, Keith J.
; APPLICANT: Johnson, Gerard P.
; APPLICANT: Pincus, Steven E.
; APPLICANT: Cox, William I.
; APPLICANT: Audonnet, Jean-Christophe Francis
; APPLICANT: Gettig, Russell Robert
; TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE
; TITLE OF INVENTION: STRAIN
; NUMBER OF SEQUENCES: 467
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; ADDRESSEE: c/o William S. Frommer
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,101
FILING DATE: 01-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2740
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 324:
SEQUENCE CHARACTERISTICS:
LENGTH: 3674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-458-101-324

Alignment Scores:
Pred. No.: 51.5 Length: 3674
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 1 Gaps: 0

US-10-024-955-7 (1-213) x US-08-458-101-324 (1-3674)

Qy 196 Lysglumethrlhysvalleuala 203
Db 2188 AAGGAGATGACTAAGTTCTGGCC 2211

RESULT 9
US-08-462-778-5
Sequence 5, Application US/08462778
Patent No. 6077517
GENERAL INFORMATION:
APPLICANT: Thomas, Wayne R.
APPLICANT: Chua, Kaw-Yan
TITLE OF INVENTION: Allergenic Protein and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,778
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/031,141
FILING DATE: 12 March 1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 24 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-462-778-5

Alignment Scores:
Pred. No.: 5.19 Length: 24
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 3 Gaps: 0

US-10-024-955-7 (1-213) x US-08-462-778-5 (1-24)

Qy 173 Serileleuasprouleph 179
Db 1 TCAATTTCGATTCATTTTC 21

RESULT 10
US-08-553-336A-5
Sequence 5, Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-553-336A-5

Alignment Scores:
Pred. No.: 5.19 Length: 24
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-08-553-336A-5 (1-24)

QY 173 Serileuappproilephe 179
|||||
Db 1 TCATTTTGGATTCATTTTC 21

RESULT 11

US-08-620-969-1/C

; Sequence 1, Application US/08620969

; Patent No. 5965126

; GENERAL INFORMATION:

; APPLICANT: Pegg, Anthony E.

; APPLICANT: Gerson, Stanton L.

; TITLE OF INVENTION: Use of Mutant Alkyltransferases For Gene

; TITLE OF INVENTION: Therapy To Protect From Toxicity of Therapeutic

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: The Pennsylvania State University

; STREET: 113 Technology Center

; CITY: University Park

; STATE: PA

; COUNTRY: US

; ZIP: 16802-7000

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/620,969

; FILING DATE: 25-MAR-1996

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Monahan, Thomas J.

; REGISTRATION NUMBER: 29,835

; REFERENCE/DOCKET NUMBER: 95-1507

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (814) 865-6277

; TELEFAX: (814) 865-3591

; INFORMATION FOR SEQ. ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 30 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: oligonucleotide

; HYPOHETICAL: NO

; ANTI-SENSE: NO

; US-08-620-969-1

Alignment Scores:

Pred. No.: 6.41 Length: 30

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.29% Indels: 0

DB: 2 Gaps: 0

US-10-024-955-7 (1-213) x US-08-620-969-1 (1-30)

QY 172 LeuSerileuappproile 178
|||||

Db 26 TTGTTCATTTTGGATTCCTATA 6

RESULT 12

US-08-404-732A-3/C

; Sequence 3, Application US/08404732A

; Patent No. 5661017

; GENERAL INFORMATION:

; APPLICANT: Dunahay, Terry G.

; APPLICANT: Roessler, Paul

; APPLICANT: Jarvis, Eric
; TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE, MATERIALS
; TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY
; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: National Renewable Energy Laboratory

; STREET: 1617 Cole Boulevard

; CITY: Golden

; STATE: CO

; COUNTRY: U.S.A.

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASCII(DOS)text (*.*)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/404,732A

; FILING DATE: 15-030-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: O'Connor, Edna M.

; REGISTRATION NUMBER: 29252

; REFERENCE/DOCKET NUMBER: 94-31

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303/275-4404

; INFORMATION FOR SEQ. ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 594 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: 3'UTR

; LOCATION: 1..594

; US-08-404-732A-3

Alignment Scores:

Pred. No.: 109 Length: 594

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.29% Indels: 0

DB: 1 Gaps: 0

US-10-024-955-7 (1-213) x US-08-404-732A-3 (1-594)

QY 177 Proilepheclyvalleuser 183
|||||

Db 197 CCAATATTGTGTCTCTCCTCC 177

RESULT 13

US-09-134-001C-1944/C

; Sequence 1944, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 1944

; LENGTH: 912

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1944

Alignment Scores:

Pred. No.:	163	Length:	912
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	4	Gaps:	0

US-10-024-955-7 (1-213) x US-09-134-001C-1944 (1-912)

QY 98 11eValysAlaHisLeuLeu 104

Db 910 ATTGTAAAGCCGACTATATA 890

RESULT 14

US-08-361-467B-1/c

; Sequence 1, Application US/08361467B

; Patent No. 5633441

; GENERAL INFORMATION:

; APPLICANT: De Greef, Willy

; APPLICANT: Van Emmelo, John

; APPLICANT: De Oliveira, Dulce E.

; APPLICANT: De Souza, Maria-Helena

; APPLICANT: Van Montagu, Marc

; TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,467B

; FILING DATE: 22-DEC-1994

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/681,492

; FILING DATE: 04-APR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/EP90/01275

; FILING DATE: 01-AUG-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 89 402 224.3

; FILING DATE: 04-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Schulman, Robert M.

; REGISTRATION NUMBER: 31,196

; REFERENCE/DOCKET NUMBER: 010830-027

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 963 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; IMMEDIATE SOURCE:

; CLONE: pmg07

; US-08-361-467B-1

Alignment Scores:

Pred. No.:

172

Length:

963

Score:

7.00

Matches:

7

Percent Similarity:

100.00%

Conservative:

0

Best Local Similarity:

100.00%

Mismatches:

0

Query Match:

3.29%

Indels:

0

DB:

1

Gaps:

0

US-10-024-955-7 (1-213) x US-08-361-467B-1 (1-963)

QY 29 11easulysAla11leaspasp 35

Db 420 ATAAACAGCCATGACGAT 400

RESULT 15

US-08-484-332C-1/c

; Sequence 1, Application US/08484332C

; Patent No. 5767374

; GENERAL INFORMATION:

; APPLICANT: De Greef, Willy

; APPLICANT: Van Emmelo, John

; APPLICANT: De Oliveira, Dulce E.

; APPLICANT: De Souza, Maria-Helena

; APPLICANT: Van Montagu, Marc

; TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,332C

; FILING DATE: 7-JUNE-1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/361,467

; FILING DATE: 22-DEC-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/681,492

; FILING DATE: 04-APR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/EP90/01275

; FILING DATE: 01-AUG-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 89 402 224.3

; FILING DATE: 04-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Schulman, Robert M.

; REGISTRATION NUMBER: 31,196

; REFERENCE/DOCKET NUMBER: 010830-093

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 963 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; IMMEDIATE SOURCE:

; CLONE: pmg07

; US-08-484-332C-1

Alignment Scores:

Pred. No.: 172 Length: 963
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 1 Gaps: 0

US-10-024-955-7 (1-213) x US-08-484-332C-1 (1-963)

OY 29 ILeasnylsAlaIleasasp 35
|||||
DB 420 ATAAACAGCCATTCACGAT 400

RESULT 16

US-08-948-564-5/C
; Sequence 5, Application US/08948564
; Patent No. 6121512
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1644 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..1542
; US-08-948-564-5

Alignment Scores:
Pred. No.: 285 Length: 1644
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 3 Gaps: 0

US-10-024-955-7 (1-213) x US-08-948-564-5 (1-1644)

OY 177 ProIlePheGlyValIleuser 183
|||||
DB 1242 CCAATTTTGGGCTCCTCAGC 1222

RESULT 17

US-08-453-848-14/C
; Sequence 14, Application US/08453848
; Patent No. 5858368
; GENERAL INFORMATION:

APPLICANT: Smith, Gale Eugene
APPLICANT: Voikovitz, Franklin
APPLICANT: Wilkinson, Bethanie Eident
APPLICANT: Voznesensky, Andrei I.
APPLICANT: Hackelt, Craig Stanway
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,848
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/120,607
; FILING DATE: 13-SEPT-1993
; CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MGS101CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-873-8794
; TELEFAX: (404)-873-8795

INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1757 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

ORGANISM: Influenza virus
INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA
FEATURE:
; NAME/KEY: polyhedrin mRNA leader (partial)
; LOCATION: 1 to 18

FEATURE:
; NAME/KEY: coding region for AcNPV 61k protein signal
; LOCATION: 19 to 72

FEATURE:
; NAME/KEY: SmaI restriction site
; LOCATION: 76 to 81

FEATURE:
; NAME/KEY: coding region for mature rHA
; LOCATION: 73 to 1728

FEATURE:
; NAME/KEY: KpnI restriction site
; LOCATION: 1735 to 1740

FEATURE:
; NAME/KEY: BglII restriction site
; LOCATION: 1741 to 1746
; FEATURE:

NAME/KEY: universal translation termination signal
LOCATION: 1747 to 1757
US-08-453-848-14

Alignment Scores:

Pred. No.:	304	Length:	1757
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	2	Gaps:	0

US-10-024-955-7 (1-213) x US-08-453-848-14 (1-1757)

Oy 175 LeuaspProIIephcglyval 181.
Db 740 CTAGACCCGATATTCGGGTT 720

RESULT 18
US-09-169-027-14/c
Sequence 14, Application US/09169027
Patent No. 6245532
GENERAL INFORMATION:
APPLICANT: Smith, Gale Eugene
APPLICANT: Volvovitz, Franklin
APPLICANT: Wilkinson, Bethanie Eident
APPLICANT: Voznesensky, Andrei I.
APPLICANT: Hackett, Craig Stanway
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/169,027
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/453,848
FILING DATE: 30-MAY-1995
APPLICATION NUMBER: 08/120,607
FILING DATE: 13-SEPT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MGS101CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1757 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Influenza virus
INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA
FEATURE:

NAME/KEY: polyhedrin mRNA leader (partial)
LOCATION: 1 to 18
FEATURE:
NAME/KEY: coding region for AcNPV 61K protein signal
NAME/KEY: sequence
LOCATION: 19 to 72
FEATURE:
NAME/KEY: SmaI restriction site
LOCATION: 76 to 81
FEATURE:
NAME/KEY: coding region for mature rHA
LOCATION: 73 to 1728
FEATURE:
NAME/KEY: KpnI restriction site
LOCATION: 1735 to 1740
FEATURE:
NAME/KEY: BglII restriction site
LOCATION: 1741 to 1746
FEATURE:
NAME/KEY: universal translation termination signal
LOCATION: 1747 to 1757
US-09-169-027-14

Alignment Scores:

Pred. No.:	304	Length:	1757
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	4	Gaps:	0

US-10-024-955-7 (1-213) x US-09-169-027-14 (1-1757)

Oy 175 LeuaspProIIephcglyval 181
Db 740 CTAGACCCGATATTCGGGTT 720

RESULT 19
US-08-453-848-6/c
Sequence 6, Application US/08453848
Patent No. 5858368
GENERAL INFORMATION:
APPLICANT: Smith, Gale Eugene
APPLICANT: Volvovitz, Franklin
APPLICANT: Wilkinson, Bethanie Eident
APPLICANT: Voznesensky, Andrei I.
APPLICANT: Hackett, Craig Stanway
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,848
FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/120,607
FILING DATE: 13-SEPT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MGS101CIP
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1793 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Influenza virus
INDIVIDUAL ISOLATE: A/Beijing/32/92 rHA
FEATURE:
NAME/KEY: polyhedrin mRNA leader (partial)
LOCATION: 1 to 18
FEATURE:
NAME/KEY: coding region for AcNPV 61k protein signal
NAME/KEY: sequence
LOCATION: 19 to 72
FEATURE:
NAME/KEY: SmaI restriction site
LOCATION: 76 to 81
FEATURE:
NAME/KEY: coding region for mature rHA
LOCATION: 73 to 1728
FEATURE:
NAME/KEY: KpnI restriction site
LOCATION: 1771 to 1777
FEATURE:
NAME/KEY: BglII restriction site
LOCATION: 1776 to 1782
FEATURE:
NAME/KEY: universal translation termination signal
LOCATION: 1783 to 1793
US-08-453-848-6

Alignment Scores:
Pred. No.: 310 Length: 1793
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 2 Gaps: 0

US-10-024-955-7 (1-213) x US-08-453-848-6 (1-1793)
QY 175 LeuAspProIlePhleglyval 181
|||||
Db 737 CTAGACCGCATATTCGGGTT 717

RESULT 20
US-09-169-027-6/C
Sequence 6, Application US/09169027
Patent No. 6245532
GENERAL INFORMATION:
APPLICANT: Smith, Gale Eugene
APPLICANT: Voliovitz, Franklin
APPLICANT: Wilkinson, Bethanie Eident
APPLICANT: Voznesensky, Andrei I.
APPLICANT: Hackett, Craig Stanway
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta

STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/169,027
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/453,848
FILING DATE: 30-MAY-1995
APPLICATION NUMBER: 08/120,607
FILING DATE: 13-SEPT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MGS101CIP
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1793 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Influenza virus
INDIVIDUAL ISOLATE: A/Beijing/32/92 rHA
FEATURE:
NAME/KEY: polyhedrin mRNA leader (partial)
LOCATION: 1 to 18
FEATURE:
NAME/KEY: coding region for AcNPV 61k protein signal
NAME/KEY: sequence
LOCATION: 19 to 72
FEATURE:
NAME/KEY: SmaI restriction site
LOCATION: 76 to 81
FEATURE:
NAME/KEY: coding region for mature rHA
LOCATION: 73 to 1728
FEATURE:
NAME/KEY: KpnI restriction site
LOCATION: 1771 to 1777
FEATURE:
NAME/KEY: BglII restriction site
LOCATION: 1776 to 1782
FEATURE:
NAME/KEY: universal translation termination signal
LOCATION: 1783 to 1793
US-09-169-027-6
Alignment Scores:
Pred. No.: 310 Length: 1793
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 4 Gaps: 0
US-10-024-955-7 (1-213) x US-09-169-027-6 (1-1793)
QY 175 LeuAspProIlePhleglyval 181
|||||
Db 737 CTAGACCGCATATTCGGGTT 717

RESULT 21

US-08-404-732A-6/c

; Sequence 6, Application US/08404732A

; Patent No. 5661017

; GENERAL INFORMATION:

; APPLICANT: Dunahay, Terry G.

; APPLICANT: Roessler, Paul

; APPLICANT: Jarvis, Eric

; TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE MATERIALS

; TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: National Renewable Energy Laboratory

; STREET: 1617 Cole Boulevard

; CITY: Golden

; STATE: CO

; COUNTRY: U.S.A.

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASCII(DOS)text (*.*)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/404,732A

; FILING DATE: 15-030-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: O'Connor, Edna M.

; REGISTRATION NUMBER: 29252

; REFERENCE/DOCKET NUMBER: 94-31

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303/275-4404

; TELEFAX: 303/275-4400

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1853 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: 1. 5'UTR

; LOCATION: 1..445

; OTHER INFORMATION: /label= ACCase

; NAME/KEY: CDS

; LOCATION: 446..1240

; OTHER INFORMATION: /label= NPRTII

; NAME/KEY: Linker Sequence

; LOCATION: 1241..1243

; NAME/KEY: ACCase 3' coding region (in non-translatable reading

; NAME/KEY: frame)

; LOCATION: 1244..1259

; NAME/KEY: 3'UTR

; LOCATION: 1260..1853

; OTHER INFORMATION: /label= ACCase

; US-08-404-732A-6

Alignment Scores:

Pred. No.:	320	Length:	1853
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	1	Gaps:	0

US-10-024-955-7 (1-213) x US-08-404-732A-6 (1-1853)

Oy 177 Proilepneglyvalleuser 183

Db 1456 CCAATATTGCTCTCTCC 1436

RESULT 22

US-08-404-732A-4/c

; Sequence 4, Application US/08404732A

; Patent No. 5661017

; GENERAL INFORMATION:

; APPLICANT: Dunahay, Terry G.

; APPLICANT: Roessler, Paul

; APPLICANT: Jarvis, Eric

; TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE MATERIALS

; TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: National Renewable Energy Laboratory

; STREET: 1617 Cole Boulevard

; CITY: Golden

; STATE: CO

; COUNTRY: U.S.A.

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASCII(DOS)text (*.*)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/404,732A

; FILING DATE: 15-030-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: O'Connor, Edna M.

; REGISTRATION NUMBER: 29252

; REFERENCE/DOCKET NUMBER: 94-31

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303/275-4404

; TELEFAX: 303/275-4400

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2120 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: 1. 5'UTR

; LOCATION: 1..445

; OTHER INFORMATION: /label= ACCase

; NAME/KEY: CDS

; LOCATION: 446..1240

; OTHER INFORMATION: /label= NPRTII

; NAME/KEY: Linker Sequence

; LOCATION: 1241..1248

; NAME/KEY: ACCase 3' coding region (in non-translatable reading

; NAME/KEY: frame)

; LOCATION: 1249..1526

; NAME/KEY: 3'UTR

; LOCATION: 1527..2120

; OTHER INFORMATION: /label= ACCase

; US-08-404-732A-4

Alignment Scores:

Pred. No.:	363	Length:	2120
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	1	Gaps:	0

US-10-024-955-7 (1-213) x US-08-404-732A-4 (1-2120)

Oy 177 Proilepneglyvalleuser 183

Db 1723 CCAATATTGCTCTCTCC 1703

Fri Feb 21 10:04:00 2003

```

CORRESPONDENCE ADDRESS:
ADDRESS: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874.186
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/782,482
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Stephen A.
REGISTRATION NUMBER: 38,609
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4848
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 2799 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: Intron
LOCATION: 1..764
FEATURE:
NAME/KEY: exon
LOCATION: 765..892
FEATURE:
NAME/KEY: Intron
LOCATION: 893..2799
US-08-874-186-42

Alignment Scores:
Pred. No.: 472 Length: 2799
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
Gaps: 0
DB: 2

US-10-024-955-7 (1-213) x US-08-874-186-42 (1-2799)
Qy 138 phevalvalalaleuSerleu 144
Db 1835 TTGTGAGTACCTAAGCCTA 1815

RESULT 26
US-09-449-285A-3/c
Sequence 3, Application US/09449285A
Patent No. 6313280
GENERAL INFORMATION:
APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie
TITLE OF INVENTION: SMAD-INTERACTING POLYPEPTIDES AND THEIR USE
FILE REFERENCE: 2676-4232US
CURRENT APPLICATION NUMBER: US/09/449,285A
CURRENT FILING DATE: 1999-11-24
PRIOR APPLICATION NUMBER: PCT/EP98/03193
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 97201645.5

PRIOR FILING DATE: 1997-06-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Patentin version 3.0
SEQ ID NO 3
LENGTH: 2959
TYPE: DNA
ORGANISM: Mus musculus
US-09-449-285A-3

Alignment Scores:
Pred. No.: 498 Length: 2959
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
Gaps: 0
DB: 4

US-10-024-955-7 (1-213) x US-09-449-285A-3 (1-2959)
Qy 8 AlaAlaValAlaPheValAla 14
Db 1823 GCAGCGTAGCCTTTGTCG 1803

RESULT 27
US-08-309-512-2
Sequence 2, Application US/08309512
Patent No. 5759828
GENERAL INFORMATION:
APPLICANT: Tal, Ronny
APPLICANT: Benziman, Moshe
APPLICANT: Gelfand, David H.
APPLICANT: Ben-Bassat, Arie
APPLICANT: Calhoon, Roger D.
APPLICANT: Wong, Hing C.
TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: U.S.A.
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/309,512
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: 8145-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 854-3660
TELEFAX: (415) 854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4558 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
```

ORGANISM: Acetobacter xylinum
US-08-309-512-2
Alignment Scores:
Pred. No.: 750 Length: 4558
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 1 Gaps: 0
US-10-024-955-7 (1-213) x US-08-309-512-2 (1-4558)
Qy 39 AAlaIeGlInSeRgIuThr 45
Db 1006 GCCATCGACGACGACGACG 1026
RESULT 28
PCT-US92-08756A-2
Sequence 2, Application PC/TUS9208756A
GENERAL INFORMATION:
APPLICANT: Tal, Ronny
APPLICANT: Benzman, Moshe
APPLICANT: Gelfand, David H.
APPLICANT: Ben-Bassat, Arle
APPLICANT: Calhoon, Roger D.
APPLICANT: Wong, Hing C.
TITLE OF INVENTION: CYCLIC DIGUANYLATE
TITLE OF INVENTION: METABOLIC ENZYMES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Limbach and Limbach
STREET: 2001 Ferry Building
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0,
SOFTWARE: Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08756A
FILING DATE: 19921014
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: WEYR 20050 USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-433-4150
TELEFAX: 415-433-8716
TELEX: 278356
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4558 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acetobacter xylinum
PCT-US92-08756A-2
Alignment Scores:
Pred. No.: 750 Length: 4558

Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 5 Gaps: 0
US-10-024-955-7 (1-213) x PCT-US92-08756A-2 (1-4558)
Qy 39 AAlaIeGlInSeRgIuThr 45
Db 1006 GCCATCGACGACGACGACG 1026
RESULT 29
US-09-120-663-1/c
Sequence 1, Application US/09120663
Patent No. 6228644
GENERAL INFORMATION:
APPLICANT: Bogdanove, Adam J.
APPLICANT: Kim, Jihyun Francis
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
TITLE OF INVENTION: ERWINIA AMILOVORA, ITS USE, AND ENCODING GENE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120,663
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/055,105
FILING DATE: 06-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1661
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5517 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-120-663-1
Alignment Scores:
Pred. No.: 899 Length: 5517
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 4 Gaps: 0
US-10-024-955-7 (1-213) x US-09-120-663-1 (1-5517)
Qy 138 PheYaiYalAlaIeuSerIeu 144
Db 4778 TTCCTGTGGCGCTATCCTTA 4758


```

? CITY: Rockville
? STATE: Maryland
? COUNTRY: USA
? ZIP: 20850
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
? COMPUTER: HP Vectra 486/33
? OPERATING SYSTEM: MSDOS version 6.2
? SOFTWARE: ASCII text
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/961.527
? FILING DATE:
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Brookes, A. Anders
? REGISTRATION NUMBER: 36,373
? REFERENCE/DOCKET NUMBER: PB340P1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (301) 309-8504
? TELEFAX: (301) 309-8512
? INFORMATION FOR SRO ID NO: 94:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8195 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
?
? US-08-961-527-94
?
?
?
? Alignment Scores:
? Pred. No.: 1.31e+03 Length: 8195
? Score: 7.00 Matches: 7
? Percent Similarity: 100.00% Conservative: 0
? Best Local Similarity: 100.00% Mismatches: 0
? Query Match: 3.25% Indels: 0
? DB: 4 Gaps: 0
?
? US-10-024-955-7 (1-213) x US-08-961-527-94 (1-8195)
?
? QY 34 AspspaatatttAaAaattt 40
? |||||||t|||||t|||||t
?
? DB 949 GATGACGCTATCGTCGCAATC 969
?
? RESULT 32
? US-08-920-422-17/c
? Sequence 17, Application US/08920422A
? Patent No. 6253473
? GENERAL INFORMATION:
? APPLICANT: Vittek, Michael P.
? APPLICANT: Mitsuda, No. 62554731aki
? APPLICANT: Roses, Allen D.
? TITLE OF INVENTION: Presentin-1 Gene Promoter
? FILE REFERENCE: VITEKPRESENTIN
? CURRENT APPLICATION NUMBER: US/08/920.422A
? CURRENT FILING DATE: 1997-08-29
? NUMBER OF SEQ ID NOS: 22
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 17
? LENGTH: 48974
? TYPE: DNA
? ORGANISM: Mus musculus
?
? US-08-920-422-17
?
? Alignment Scores:
? Pred. No.: 7.12e+03 Length: 48974
? Score: 7.00 Matches: 7
? Percent Similarity: 100.00% Conservative: 0
? Best Local Similarity: 100.00% Mismatches: 0
? Query Match: 3.25% Indels: 0
? DB: 4 Gaps: 0

```

US-10-024-955-7 (1-213) x US-08-920-422-17 (1-48974)
QY 119 AlatyrlsleugllyspIeu 125
Db 28905 GCTTACAGCTTGAGATTGTG 28885

RESULT 33
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Alignment Scores:
Pred. No.: 3.84e+05 Length: 4403765
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-09-103-840A-2 (1-4403765)
QY 33 lIeaspsAlaIleAlaIa 39
Db 227117 ATAGACGACGCGATCGCGCC 227137

RESULT 34
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 3.85e+05 Length: 4411529
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-09-103-840A-1 (1-4411529)
QY 33 lIeaspsAlaIleAlaIa 39
Db 227008 ATAGACGACGCGATCGCGCC 227028

RESULT 35
US-08-462-778-3
; Sequence 3, Application US/08462778
; Patent No. 6077517
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kaw-Yan
; TITLE OF INVENTION: Allergenic Protein and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lohive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,778
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/031,141
; FILING DATE: 12 March 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-053CP (IMF-032CP)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-462-778-3

Alignment Scores:
Pred. No.: 46.9 Length: 18
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
DB: 3 Gaps: 0

US-10-024-955-7 (1-213) x US-08-462-778-3 (1-18)
QY 18 AspProIleHisTyraSp 23
Db 1 GATCCATTCACATGAT 18

RESULT 36
US-08-553-336A-3
; Sequence 3, Application US/08553336A
; Patent No. 6413738

```

; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-553-336A-3

Alignment Scores:
Pred. No.: 46.9 Length: 18
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
Gaps: 0
DB: 4

US-10-024-955-7 (1-213) x US-08-553-336A-3 (1-18)
QY 18 Aspprollehtytrasp 23
Db 1 GATCCAAATTCAGTATGAT 18

RESULT 37
US-08-553-336A-15/c
; Sequence 15, Application US/08553336A
; Patent No. 6413738
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-553-336A-15

Alignment Scores:
Pred. No.: 46.9 Length: 18
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
Gaps: 0
DB: 4

US-10-024-955-7 (1-213) x US-08-553-336A-15 (1-18)
QY 56 Asplysphegluarghis 61
Db 18 GATAAATTCGAGCGTCAT 1

RESULT 38
US-09-297-535-4/c
; Sequence 4, Application US/09297535
; Patent No. 6268199
; GENERAL INFORMATION:
; APPLICANT: Meulenbergh, Johanna J.M.
; APPLICANT: Pol, Johannes M.A.
; APPLICANT: Bos-de Ruijter, Judy N.A.
; TITLE OF INVENTION: Infectious clones of RNA viruses and vaccines and
; FILE REFERENCE: P20749US00
; CURRENT APPLICATION NUMBER: US/09/297,535
; CURRENT FILING DATE: 1999-10-12
; EARLIER APPLICATION NUMBER: PCT/NL97/00593
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: 96203024.3
; EARLIER FILING DATE: 1996-10-30
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 20
; TYPE: DNA
; ORGANISM: LeJystad virus
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: (1)..(20)
; OTHER INFORMATION: /note="Primer ALG4"
US-09-297-535-4

Alignment Scores:
Pred. No.: 51.8 Length: 20
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 2.82% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-09-297-535-4 (1-20)

QY 172 LeuSer11eLeuAspPro 177
DB 20 CTATCGATCTCGATCCT 3

RESULT 39

US-08-479-852-43
Sequence 43, Application US/08479852

Patent No. 5712385

GENERAL INFORMATION:
APPLICANT: Sherrol H. McDonough, Thomas B. Ryder,

APPLICANT: Yeasing Yang

TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION

TITLE OF INVENTION: OLIGONUCLEOTIDES AND PROBES

NUMBER OF SEQUENCES: 139

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon

STREET: 611 West Sixth Street

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90017

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

COMPUTER: IBM PS/2 Model 502 or 555X

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: Wordperfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,852

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/040,745

FILING DATE:

APPLICATION NUMBER: U.S. Serial No. 5712385 07/550,837

FILING DATE: 7/10/90

APPLICATION NUMBER: U.S. Serial No. 5712385 07/379,501

FILING DATE: 7/11/89

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 196/189

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 21

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-479-852-43

Alignment Scores:

Pred. No.: 54.3 Length: 21

Score: 6.00 Matches: 6

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.82% Indels: 0

DB: 1 Gaps: 0

US-10-024-955-7 (1-213) x US-08-479-852-43 (1-21)

QY 12 PheValAlaValSerAla 17

DB 2 TTGCTGCTGCTGCTGCT 19

RESULT 40

US-08-479-852-123
Sequence 123, Application US/08479852

QY 172 LeuSer11eLeuAspPro 177
DB 20 CTATCGATCTCGATCCT 3

RESULT 41

US-08-462-646-43
Sequence 43, Application US/08462646

Patent No. 5856088

GENERAL INFORMATION:
APPLICANT: Sherrol H. McDonough, Thomas B. Ryder,

APPLICANT: Yeasing Yang

TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION

TITLE OF INVENTION: OLIGONUCLEOTIDES AND PROBES

NUMBER OF SEQUENCES: 139

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon

STREET: 611 West Sixth Street

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90017

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

COMPUTER: IBM PS/2 Model 502 or 555X

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: Wordperfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,852

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/040,745

FILING DATE:

APPLICATION NUMBER: U.S. Serial No. 5712385 07/550,837

FILING DATE: 7/10/90

APPLICATION NUMBER: U.S. Serial No. 5712385 07/379,501

FILING DATE: 7/11/89

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 196/189

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 123:

SEQUENCE CHARACTERISTICS:

LENGTH: 21

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-479-852-123

Alignment Scores:

Pred. No.: 54.3 Length: 21

Score: 6.00 Matches: 6

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.82% Indels: 0

DB: 1 Gaps: 0

US-10-024-955-7 (1-213) x US-08-479-852-123 (1-21)

QY 12 PheValAlaValSerAla 17

DB 2 UUCGUCGUCGUCUCGCU 19

TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,646
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/040,745
FILING DATE: 26-MAR-1993
APPLICATION NUMBER: U.S. Serial No. 5856088 07/550,837
FILING DATE: 7/10/90
APPLICATION NUMBER: U.S. Serial No. 5856088 07/379,501
FILING DATE: 7/11/89
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 196/189
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 21
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-462-646-43
Alignment Scores:
Pred. No.: 54.3 Length: 21
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
Gaps: 0
US-10-024-955-7 (1-213) x US-08-462-646-43 (1-21)
QY 12 PheValAlaValSerAla 17
Db 2 TTCGTCGCTGCTCCGCT 19
RESULT 42
US-08-462-646-123
Sequence 123, Application US/08462646
Patent No. 5856088
GENERAL INFORMATION:
APPLICANT: Sherrol H. McDonough, Thomas B. Ryder,
APPLICANT: Yeasing Yang
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION
TITLE OF INVENTION: OLIGONUCLEOTIDES AND PROBES
NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,646
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/040,745
FILING DATE: 26-MAR-1993
APPLICATION NUMBER: U.S. Serial No. 5856088 07/550,837
FILING DATE: 7/10/90
APPLICATION NUMBER: U.S. Serial No. 5856088 07/379,501
FILING DATE: 7/11/89
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 196/189
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 123:
SEQUENCE CHARACTERISTICS:
LENGTH: 21
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-462-646-123
Alignment Scores:
Pred. No.: 54.3 Length: 21
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
Gaps: 0
US-10-024-955-7 (1-213) x US-08-462-646-123 (1-21)
QY 12 PheValAlaValSerAla 17
Db 2 UUCGUCGUCGUCUCCGCU 19
RESULT 43
US-09-013-406-43
Sequence 43, Application US/09013406
Patent No. 6252059
GENERAL INFORMATION:
APPLICANT: Sherrol H. McDonough, Thomas B. Ryder,
APPLICANT: Yeasing Yang
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION
TITLE OF INVENTION: OLIGONUCLEOTIDES AND PROBES
NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/013,406
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,852
FILING DATE:
APPLICATION NUMBER: U.S. Serial No. 6252059 07/550,837
FILING DATE: 7/10/90
APPLICATION NUMBER: U.S. Serial No. 6252059 07/379,501
FILING DATE: 7/11/89
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 196/189
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 21
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-013-406-43

Alignment Scores:
Pred. No.: 54.3 Length: 21
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-09-013-406-43 (1-21)

QY 12 Phevalalavalserala 17
|||||
DB 2 TTCGTCGCTGTCTCCGCT 19

RESULT 44
US-09-013-406-123
Sequence 123, Application US/09013406
Patent No. 6252059
GENERAL INFORMATION:
APPLICANT: Sherrol H. McDonough, Thomas B. Rydér,
APPLICANT: Yeasling Yang
TITLE OF INVENTION: NUCLEIC ACID AMPLIFICATION
TITLE OF INVENTION: OLIGONUCLEOTIDES AND PROBES
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB storage
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/013,406
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,852
FILING DATE:
APPLICATION NUMBER: U.S. Serial No. 6252059 07/550,837
FILING DATE: 7/10/90
APPLICATION NUMBER: U.S. Serial No. 6252059 07/379,501
FILING DATE: 7/11/89
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 196/189
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 123:
SEQUENCE CHARACTERISTICS:
LENGTH: 21
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-013-406-123

Alignment Scores:
Pred. No.: 54.3 Length: 21
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-09-013-406-123 (1-21)

QY 12 Phevalalavalserala 17
|||||
DB 2 UUCGUCGUCGUCUCCGCU 19

RESULT 45
US-08-379-078-654
Sequence 654, Application US/08379078
Patent No. 5639612
GENERAL INFORMATION:
APPLICANT: Mitsunashi, Masato
APPLICANT: Cooper, Allan
TITLE OF INVENTION: Gene Detection System
NUMBER OF SEQUENCES: 726
CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,078
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/974,406
FILING DATE: 12-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: HITACHI.011CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
INFORMATION FOR SEQ ID NO: 654:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-379-078-654

Alignment Scores:

Pred. No.: 56.7 Length: 22
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
DB: 1 Gaps: 0

US-10-024-955-7 (1-213) x US-08-379-078-654 (1-22)

QY 100 LysAlaHisLeuLeu1le 105

DB 5 AAAGCCACCTGCTCATY 22

RESULT 46

US-07-974-409C-273

Sequence 273, Application US/07974409C

Patent No. 6300058

GENERAL INFORMATION:

APPLICANT: Akiyaya, Tatsuo

APPLICANT: Mitsubashi, Masato

APPLICANT: Cooper, Allen

TITLE OF INVENTION: METHOD AND REAGENT

NUMBER OF SEQUENCES: 457

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb, Martens, Olson, and Bear

STREET: 620 Newport Center Dr. Sixteenth Floor

CITY: Newport Beach

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/974,409C

FILING DATE: 12-NOV-1992

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Altman, Daniel E.

REGISTRATION NUMBER: 34,115

REFERENCE/DOCKET NUMBER: HITACHI.006CP2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-9502

TELEFAX: 714-760-0404

INFORMATION FOR SEQ ID NO: 273:

SEQUENCE CHARACTERISTICS:

LENGTH: 22

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-07-974-409C-273

Alignment Scores:

Pred. No.: 56.7 Length: 22
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-07-974-409C-273 (1-22)

QY 100 LysAlaHisLeuLeu1le 105

DB 5 AAAGCCACCTGCTCATY 22

RESULT 47

PCT-US93-00977-273

Sequence 273, Application PC/TUS9300977

GENERAL INFORMATION:

TITLE OF INVENTION: METHOD AND REAGENT FOR MEASURING MESSENGER RNA

NUMBER OF SEQUENCES: 711

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb, Martens, Olson, and Bear

STREET: 620 Newport Center Dr. Sixteenth Floor

CITY: Newport Beach

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/00977

FILING DATE: 19930129

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Altman, Daniel E.

REGISTRATION NUMBER: 34,115

REFERENCE/DOCKET NUMBER: HITACHI.006H

TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-9502

TELEFAX: 714-760-0404

INFORMATION FOR SEQ ID NO: 273:

SEQUENCE CHARACTERISTICS:

LENGTH: 22

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

PCT-US93-00977-273

Alignment Scores:

Pred. No.: 56.7 Length: 22
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.82% Indels: 0
DB: 5 Gaps: 0

US-10-024-955-7 (1-213) x PCT-US93-00977-273 (1-22)

QY 100 LysAlaHisLeuLeu1le 105

DB 5 AAAGCCACCTGCTCATY 22

RESULT 48

US-09-068-880-17

Sequence 17, Application US/09068880B

Patent No. 6203962

GENERAL INFORMATION:

APPLICANT: Nunokawa, Youichi

APPLICANT: Oikawa, Shinzo

APPLICANT: Tanaka, Shoji

TITLE OF INVENTION: Method for Screening Compounds

TITLE OF INVENTION: Regulating the Expression of Human-Inducible Nitric Oxide

FILE REFERENCE: SHIM-001

CURRENT FILING DATE: 1998-09-02

EARLIER APPLICATION NUMBER: PCT/JP97/03303

CURRENT FILING DATE: 1997-09-18

NUMBER OF SEQ ID NOS: 17

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 17
LENGTH: 26
TYPE: DNA
ORGANISM: Homo sapiens
US-09-068-880-17

Alignment Scores:

Pred. No.:	66.4	Length:	26
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.82%	Indels:	0
DB:	4	Gaps:	0

US-10-024-955-7 (1-213) x US-09-068-880-17 (1-26)

QY 142 LeuSerLeuGluLeuSer 147
|||||

Db 6 CTGCTCTGGAAATTTCT 23

RESULT 49

US-08-553-336A-8
Sequence 8, Application US/08553336A

Patent No. 6413738

GENERAL INFORMATION:

APPLICANT: Wayne R. Thomas and Kaw-Yan Chua

TITLE OF INVENTION: Allergenic Proteins and Peptides From

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/553,336A

FILING DATE: 10-JUN-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/081,540

FILING DATE: 22-JUNE-1993

ATTORNEY/AGENT INFORMATION:

NAME: Jane E. Remillard

REGISTRATION NUMBER: 38,872

REFERENCE/DOCKET NUMBER: IMI-032CP2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-553-336A-8

Alignment Scores:

Pred. No.:	66.4	Length:	26
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.82%	Indels:	0
DB:	4	Gaps:	0

US-10-024-955-7 (1-213) x US-08-553-336A-8 (1-26)

QY 18 AspProIleHisTyrAsp 23
|||||

Db 9 GATCAATTCACATATGAT 26

RESULT 50

US-08-279-751B-17
Sequence 17, Application US/08279751B

Patent No. 5629413

GENERAL INFORMATION:

APPLICANT: Todd C. Peterson

TITLE OF INVENTION: OLIGONUCLEOTIDES WITH

TITLE OF INVENTION: ACTIVITY AGAINST HUMAN

TITLE OF INVENTION: IMMUNODEFICIENCY

NUMBER OF SEQUENCES: 95

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon

STREET: Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

MEDIUM TYPE: storage

COMPUTER: IBM compatible

OPERATING SYSTEM: IBM P.C. DOS 6.0

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/279,751B

FILING DATE: July 19, 1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/094,390

FILING DATE: July 19, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Hebert, Sheldon O.

REGISTRATION NUMBER: 38,179

REFERENCE/DOCKET NUMBER: 208/075

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 27 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-279-751B-17

Alignment Scores:

Pred. No.:	68.8	Length:	27
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.82%	Indels:	0
DB:	1	Gaps:	0

US-10-024-955-7 (1-213) x US-08-279-751B-17 (1-27)

QY 12 PheValAlaValSerAla 17
|||||

Db 7 TTGCTGCTGTCTCCGCT 24

Search completed: February 20, 2003, 19:56:50
Job time: 1897 secs

GenCore version 5.1.4.P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 19:07:15 : Search time 113 Seconds
(Without alignments)
960.053 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213
Sequence: 1 MMKFLIAAFAVAVSADPI.....VRKEMTKVLAPFRKRELEKN 213

Scoring table:

OLIGO	Xgapop 60.0	Xgapext 60.0
	Xgapop 60.0	Xgapext 60.0
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 424239 seqs, 254661826 residues

Word size: 1

Total number of hits satisfying chosen parameters: 847233

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL-frame+ .p2n.model -DEV-xlp
-O=/cgn2_1/USPTO.spool/US10024955/runat.14022003_130738_24369/app-query.fasta.1.391
-DB=Published_Applications.NA -QFMT=fastap -SUFFIX=inp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS-human40.cdt -LIST=45 -DOCAUGEN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=50 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10024955@cgn.1.1.80 @runat.14022003_130738_24369
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : Published Applications.NA.*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/1/pubpna/PTCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/1/pubpna/PCITUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	761	9	US-10-024-955-6
2	25	11.7	812	9	US-10-024-955-1
3	9	4.2	317	9	US-09-736-457-1046
4	4	4.2	317	9	US-09-902-941-1046

C	5	4.2	317	9	US-09-849-626-1046	Sequence 1046, Ap
C	6	3.8	24	9	US-10-024-955-12	Sequence 12, Appl
C	7	3.8	1128	9	US-09-738-626-3475	Sequence 3475, Ap
C	8	3.8	6027	10	US-09-070-927A-124	Sequence 124, App
C	9	3.3	24	9	US-10-024-955-5	Sequence 5, Appl
C	10	3.3	122	10	US-09-560-863-686	Sequence 686, App
C	11	3.3	264	10	US-09-815-242-9182	Sequence 9182, Ap
C	12	3.3	289	10	US-09-294-093B-2975	Sequence 2975, Ap
C	13	3.3	306	10	US-09-974-300-1462	Sequence 1462, Ap
C	14	3.3	349	10	US-09-878-574-1250	Sequence 1250, Ap
C	15	3.3	352	10	US-09-878-574-1466	Sequence 1466, Ap
C	16	3.3	362	10	US-09-983-965-1072	Sequence 1072, Ap
C	17	3.3	397	10	US-09-878-574-1600	Sequence 1600, Ap
C	18	3.3	401	9	US-09-946-607-407	Sequence 407, App
C	19	3.3	401	10	US-09-795-668-407	Sequence 407, App
C	20	3.3	401	10	US-09-795-668-407	Sequence 407, App
C	21	3.3	409	10	US-09-960-352-14798	Sequence 14798, A
C	22	3.3	471	10	US-09-867-701-6524	Sequence 6524, Ap
C	23	3.3	501	10	US-09-864-761-13286	Sequence 13286, A
C	24	3.3	555	9	US-10-040-739-275	Sequence 275, App
C	25	3.3	600	9	US-09-860-670-54	Sequence 54, Appl
C	26	3.3	600	9	US-09-989-442-17	Sequence 17, Appl
C	27	3.3	600	10	US-09-815-242-6072	Sequence 6072, Ap
C	28	3.3	600	10	US-09-764-853-71	Sequence 71, Appl
C	29	3.3	777	10	US-09-770-445-885	Sequence 885, App
C	30	3.3	846	10	US-09-799-777-82	Sequence 82, Appl
C	31	3.3	913	9	US-09-764-868-29	Sequence 29, Appl
C	32	3.3	924	10	US-09-815-242-7209	Sequence 7209, Ap
C	33	3.3	939	10	US-09-886-055-290	Sequence 290, App
C	34	3.3	1065	10	US-09-778-844-120	Sequence 120, App
C	35	3.3	1267	10	US-09-925-299-125	Sequence 125, App
C	36	3.3	1332	10	US-09-974-300-5126	Sequence 5126, Ap
C	37	3.3	1479	9	US-09-938-842A-932	Sequence 932, App
C	38	3.3	1511	9	US-09-989-442-64	Sequence 64, Appl
C	39	3.3	1511	10	US-09-764-853-308	Sequence 308, App
C	40	3.3	1605	9	US-09-932-367A-1	Sequence 1, Appl
C	41	3.3	1658	9	US-09-932-367A-13	Sequence 13, Appl
C	42	3.3	1664	9	US-09-932-367A-15	Sequence 15, Appl
C	43	3.3	1782	9	US-10-116-252-2	Sequence 2, Appl
C	44	3.3	1902	9	US-09-938-842A-2865	Sequence 2865, Ap
C	45	3.3	2103	10	US-09-733-757-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-024-955-6
Sequence 6, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From House Dust Mite and Uses Thereof
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996

APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMT-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 761 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 43..681
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-024-955-6

Alignment Scores:
Pred. No.: 7,026-211 Length: 761
Score: 213.00 Matches: 213
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x US-10-024-955-6 (1-761)

QY 1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
DB 43 ATGATCAAAATTTTGTGTTGCTGCGCGGCGATTGTCGCCCTTCGGCTGATCCAAAT 102
QY 21 HlSTYrAsPlYsIlEThrGlUglUlLeAsnLYsAlAlEtsPAsPalAlEalaIle 40
DB 103 CACTATGATAAATCACCCAGAAATCAACAAAGCTATTGATGATGCCATTGCTCTATT 162
QY 41 GlUInSerGlnThrIleAspProMetLysValProAspHisAlaAspLysPheGluArg 60
DB 163 GAACAATCCGAAACAAATGATCCAAATGAAAGTACCTGATCATGCCGATAAATTCGAACGT 222
QY 61 HlSvAlGlylLeValAspPheLysGlyLulLeuAlaMetArgAsnIleGluAlaArgGly 80
DB 223 CATGTTGATTTGTGATTTCAAGGTGAATTAGCCATCGGAACATTCAGGCTCGAGGA 282
QY 81 LeuLYSGlnMetLysArgGlnGlnLysAlaAsnValLysGlyGluGlylLeValLys 100
DB 283 TTGAAACAATGAAACGTCAGGTGCTAAATGTCAAAGGTGAAGGGTATTGTTAA 342
QY 101 AlAlHlSLeuLeuIleGlyValHlSAspAspIleValSerMetGlnUtyrAspLeuAlaTyr 120
DB 343 GCTCATTTTGTGTCGCTTCGATGATATCTGCTCGATGGAATATGATTTAGCATAC 402
QY 121 LysLeuGlyAspLeuHisProThrHrHlSValIleSerAspIleGlnAspPheValAl 140
DB 403 AAATTCGGTGAATCTTCACCAACCACTCATGTCATTTGCGATTCGAAGTTTGTGTT 462
QY 141 AlAlSerLeuGluLlSerAspGlyLysAlIerHlMetHlSerPheGluValArg 160
DB 463 GCGTGTCCCTTGAAATTTCTGTGAAGGTAAACATGACATCTTTTGAAATGACA 522
QY 161 GlnPheAlaAsnValValAsnHlSIlleGlyLysLeuSerIleLeuAspProIlePheGly 180
DB 523 CAATTCGCTAAATGTTTCAACCATATTGGTGTTCATATCTTGATTCGAATTTTGGC 582
QY 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys 200
DB 583 GTTTATATCGATGATATTGACCGGTATTTTCCAAGACACCGTACGTAGAGAAATGACCAA 642
QY 201 ValLeuAlaProAlaPheLysArgGluLeuGlyLysAsn 213

DB 643 GTATTGGCACAGCAATTAAACGTGAATTGAAAAAAT 681
RESULT 2
US-10-024-955-1
Sequence 1, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,36A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMT-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 812 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 68..712
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 119..712
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-024-955-1
Alignment Scores:
Pred. No.: 5,396-17 Length: 812
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.74% Indels: 0
Gaps: 0
US-10-024-955-7 (1-213) x US-10-024-955-1 (1-812)
QY 155 ThrSerPheGluValArgGlnPheAlaAsnValValAsnHlSIlleGlyLysLeuSerIle 174
DB 530 ACATCGTTCGAAGTACGTCGAATTTGCCAATGTTGCATCATGTTGCTTCAATT 589
QY 175 LeuAspProIlePhe 179
DB 590 TTGATCCAAATTTTC 604

```
RESULT 3
US-09-736-457-1046/c
: Sequence 1046, Application US/09736457
: Patent No. US20020168637A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Tonglong
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: Lodes, Michael A.
: APPLICANT: Fanger, Gary
: APPLICANT: Vedvick, Tom
: APPLICANT: Carter, Darick
: APPLICANT: Retter, Marc
: APPLICANT: Mannion, Jane
: APPLICANT: Fan, Liqun
: APPLICANT: Wang, Aijun
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.478C15
: CURRENT APPLICATION NUMBER: US/09/736.457
: CURRENT FILING DATE: 2000-12-13
: NUMBER OF SEQ ID NOS: 1864
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1046
: LENGTH: 317
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-736-457-1046

Alignment Scores:
Pred. No.: 0.728 Length: 317
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x US-09-736-457-1046 (1-317)
Oy 188 AlaiIepheGInAspThrValArgLys 196
Db 309 GCCATTTCAGACGACGCGTAGAGAG 283

RESULT 4
US-09-902-941-1046/c
: Sequence 1046, Application US/09902941
: Patent No. US20020172952A1
: GENERAL INFORMATION:
: APPLICANT: Henderson, Robert A.
: APPLICANT: Wang, Tonglong
: APPLICANT: Watanabe, Yoshihiro
: APPLICANT: Johnson, Jeffrey C.
: APPLICANT: Retter, Marc W.
: APPLICANT: Marnerakis, Margarita
: APPLICANT: Carter, Darick
: APPLICANT: Fanger, Gary R.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: McNabb, Andria
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.478C17
: CURRENT APPLICATION NUMBER: US/09/902.941
: CURRENT FILING DATE: 2001-07-10
: NUMBER OF SEQ ID NOS: 2002
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1046
: LENGTH: 317
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-902-941-1046

Alignment Scores:
```

```
Pred. No.: 0.728 Length: 317
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x US-09-902-941-1046 (1-317)
Oy 188 AlaiIepheGInAspThrValArgLys 196
Db 309 GCCATTTCAGACGACGCGTAGAGAG 283

RESULT 5
US-09-849-626-1046/c
: Sequence 1046, Application US/09849626
: Publication No. US20020197659A1
: GENERAL INFORMATION:
: APPLICANT: Bangur, Chaitanya
: APPLICANT: Fanger, Gary
: APPLICANT: Wang, Aijun
: APPLICANT: Wang, Tonglong
: APPLICANT: Switzer, Anne
: APPLICANT: McNeill, Patricia
: APPLICANT: Clapper, Jonathan
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.478C16
: CURRENT APPLICATION NUMBER: US/09/849.626
: CURRENT FILING DATE: 2001-05-03
: NUMBER OF SEQ ID NOS: 1926
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1046
: LENGTH: 317
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-849-626-1046

Alignment Scores:
Pred. No.: 0.728 Length: 317
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x US-09-849-626-1046 (1-317)
Oy 188 AlaiIepheGInAspThrValArgLys 196
Db 309 GCCATTTCAGACGACGCGTAGAGAG 283

RESULT 6
US-10-024-955-12
: Sequence 12, Application US/10024955
: Patent No. US20020168373A1
: GENERAL INFORMATION:
: APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
: TITLE OF INVENTION: Allergenic Proteins and Peptides From
: TITLE OF INVENTION: House Dust Mite and Uses Therefor
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD, LLP
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
```

;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/024,955
;; FILING DATE: 19-Dec-2001
;; CLASSIFICATION: <Unknown>
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/553,336A
;; FILING DATE: 10-JUN-1996
;; APPLICATION NUMBER: US 08/081,540
;; FILING DATE: 22-JUNE-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jane E. Remillard
;; REGISTRATION NUMBER: 38,872
;; REFERENCE/DOCKET NUMBER: IMI-032CP2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)227-7400
;; TELEFAX: (617)742-4214
;; INFORMATION FOR SEQ ID NO: 12:
;;
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 24 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
;;
;; US-10-024-955-12
;;
Alignment Scores:
Pred. No.: 0.764 Length: 24
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 9 Gaps: 0
US-10-024-955-7 (1-213) x US-10-024-955-12 (1-24)
QY 173 Ser11e1euaSpPro11epegly 180
DB 1 TCAATCTTGATCCAAATTTTGGC 24
RESULT 7
US-09-738-626-3475/C
; Sequence 3475, Application US/09738626
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3475
; LENGTH: 1128
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3475

Alignment Scores:
Pred. No.: 24.6 Length: 1128
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 9 Gaps: 0
US-10-024-955-7 (1-213) x US-09-738-626-3475 (1-1128)
QY 8 Alalavala1aphevalalaval 15
DB 38 GCGCAGTTCCTTGTCTGCTGC 15
RESULT 8
US-09-070-927A-124/C
; Sequence 124, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 124:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6027 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 124:
; US-09-070-927A-124
Alignment Scores:
Pred. No.: 111 Length: 6027
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 10 Gaps: 0
US-10-024-955-7 (1-213) x US-09-070-927A-124 (1-6027)

Oy 6 LeuilealaAlaValAlaPheVal 13
|||||
DB 2072 TTAATTGACGCGTAGCATTTGTT 2049

RESULT 9
US-10-024-955-5
Sequence 5, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-024-955-5

Alignment Scores:
Pred. No.: 8.21 Length: 24
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x US-10-024-955-5 (1-24)
Oy 173 SerileuaspProilePhe 179
|||||
DB 1 TCAATTTGATCCAAATTTTC 21

RESULT 10
US-09-560-863-686/C
Sequence 686, Application US/09560863
Patent No. US20020110809A1
GENERAL INFORMATION:
APPLICANT: Nehls, Michael C.
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.

TITLE OF INVENTION: No. US20020110809A1 Human Polynucleotides and the
FILE REFERENCE: LEX-0018-USA
CURRENT APPLICATION NUMBER: US/09/560,863
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/132,408
PRIOR FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 1008
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 686
LENGTH: 122
TYPE: DNA
ORGANISM: homo sapiens
US-09-560-863-686

Alignment Scores:
Pred. No.: 35.6 Length: 122
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-560-863-686 (1-122)
Oy 93 LysGlyGluGlyLeuVal 99
|||||
DB 40 AAAGTGAGGAGCATGATGTC 20

RESULT 11
US-09-815-242-9182
Sequence 9182, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 9182
LENGTH: 264
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(264)
US-09-815-242-9182

Alignment Scores:

; SOFTWARE: FASTSEQ FOR WINDOWS VERSION 4.0

; NUMBER OF SEQ ID NOS: 15/15

SEQ ID NO 1466
LENGTH: 352
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-037-Q1-B1-E1
US-09-878-574-1466

Alignment Scores:

Pred. No.:	92.4	Length:	352
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	10	Gaps:	0

US-10-024-955-7 (1-213) x US-09-878-574-1466 (1-352)

OY 102 HlsleuLeuIleGlyValHis 108
|||||

DB 142 CATCTTCATTCATGGGCTCCAC 162

RESULT 16

US-09-983-965-1072
Sequence 1072, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengding
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 1072
LENGTH: 362
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 57-LIB188-007-Q1-E1-G10
US-09-983-965-1072

Alignment Scores:

Pred. No.:	94.8	Length:	362
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	10	Gaps:	0

US-10-024-955-7 (1-213) x US-09-983-965-1072 (1-362)

OY 14 AlavaISeraIasPProile 20
|||||

DB 230 GCAGTCAGTCAGACCCGATA 250

RESULT 17

US-09-878-574-1600
Sequence 1600, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B

CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 1600
LENGTH: 397
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-035-Q1-B1-F7
US-09-878-574-1600

Alignment Scores:

Pred. No.:	103	Length:	397
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	10	Gaps:	0

US-10-024-955-7 (1-213) x US-09-878-574-1600 (1-397)

OY 102 HlsleuLeuIleGlyValHis 108
|||||

DB 142 CATCTTCATTCATGGGCTCCAC 162

RESULT 18

US-09-946-807-407
Sequence 407, Application US/09946807
Patent No. US20020165144A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinhorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
FILE REFERENCE: 2345.2004-001
CURRENT APPLICATION NUMBER: US/09/946,807
CURRENT FILING DATE: 2001-09-05
PRIOR APPLICATION NUMBER: US/09/795,668
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/515,716
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 1531
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 407
LENGTH: 401
TYPE: DNA
ORGANISM: Homo sapiens
US-09-946-807-407

Alignment Scores:

Pred. No.:	104	Length:	401
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.29%	Indels:	0
DB:	9	Gaps:	0

US-10-024-955-7 (1-213) x US-09-946-807-407 (1-401)

OY 102 HlsleuLeuIleGlyValHis 108
|||||

DB 357 CATCTTCATTCATGGGCTTCAT 377

RESULT 19

US-09-795-668-407
Sequence 407, Application US/09795668
Patent No. US20020045577A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinhorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE

```

; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 407
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-407

Alignment Scores:
Pred. No.: 104          Length: 401
Score: 7.00           Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 3.29%          Indels: 0
DB: 10                  Gaps: 0

US-10-024-955-7 (1-213) x US-09-795-668-407 (1-401)

QY 102 HisleuleuileglyvalHis 108
DB 357 CATCTGTGATGTGGGTTTCAT 377

RESULT 20
US-09-795-686-407
; Sequence 407, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 407
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-407

Alignment Scores:
Pred. No.: 104          Length: 401
Score: 7.00           Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 3.29%          Indels: 0
DB: 10                  Gaps: 0

US-10-024-955-7 (1-213) x US-09-795-686-407 (1-401)

QY 102 HisleuleuileglyvalHis 108
DB 357 CATCTGTGATGTGGGTTTCAT 377

RESULT 21
US-09-960-352-14798
; Sequence 14798, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

```

```

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14798
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 63-LIB34-018-Q1-E1-H12
US-09-960-352-14798

Alignment Scores:
Pred. No.: 106          Length: 409
Score: 7.00           Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 3.29%          Indels: 0
DB: 10                  Gaps: 0

US-10-024-955-7 (1-213) x US-09-960-352-14798 (1-409)

QY 1 MetMetlyspheuleuile 7
DB 5 ATGATGAATTCCTGCTCAT 25

RESULT 22
US-09-867-701-6524
; Sequence 6524, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6524
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-6524

Alignment Scores:
Pred. No.: 120          Length: 471
Score: 7.00           Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 3.29%          Indels: 0
DB: 10                  Gaps: 0

US-10-024-955-7 (1-213) x US-09-867-701-6524 (1-471)

QY 177 Protlephecglyvalleuser 183
DB 65 CCCATCTTTGAGTGCCTCTCA 85

RESULT 23
US-09-864-761-13286/C
; Sequence 13286, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Neomica-X-1

```


;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 13286
;; LENGTH: 501
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL136226.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.5
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.6
US-09-864-761-13286

Alignment Scores:
Pred. No.: 127 Length: 501
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-864-761-13286 (1-501)

Oy 186 leuthralalephcglasp 192
Db 268 CTACTGCTATTTCACAGAT 248

RESULT 24
US-10-040-739-275/c
; Sequence 275, Application US/10040739
; Patent No. US20020173635A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John

;; LaVallie, Edward
;; Racie, Lisa
;; Merberg, David
;; Treacy, Maurice
;; Spaulding, Vikki
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 1519
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 275:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
SEQUENCE DESCRIPTION: SEQ ID NO: 275:
US-10-040-739-275

Alignment Scores:
Pred. No.: 139 Length: 555
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x US-10-040-739-275 (1-555)

Oy 34 aspaspalaaleaale 40
Db 414 GACGACGCATCGCGGCATC 394

RESULT 25
US-09-860-670-54/c
; Sequence 54, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA127PI
; CURRENT APPLICATION NUMBER: US/09/860,670
; CURRENT FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens

```
US-09-860-670-54
Alignment Scores:
Pred. No.: 149
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 9
Gaps: 0
US-10-024-955-7 (1-213) x US-09-860-670-54 (1-600)
QY 141 AlaLeuSerIleuGlutIleSer 147
|||||
DB 52 GCCTTATCCCTGCAATTCCT 32
RESULT 26
US-09-989-442-17/C
; Sequence 17, Application US/09989442
; Publication No. US20030013649A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0208
; CURRENT APPLICATION NUMBER: US/09/989,442
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
```

PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 60/225,266
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/232,400
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,414
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,244
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,064
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/233,063
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,401
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/241,808
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,826
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,786
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,221
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,475
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/231,243
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,065
PRIOR FILING DATE: 2000-09-14

Alignment Scores:
Pred. No.: 149
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 9
Gaps: 0

US-10-024-955-7 (1-213) x US-09-989-442-17 (1-600)

QY 141 AlaleuSerLeuGluIleSer 147
|||||
Db 52 GCTTATCCCTGGAATTTCT 32

RESULT 27
US-09-815-242-6072/c
Sequence 6072, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlson, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
CURRENT FILING DATE: 2001-03-21
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6072
LENGTH: 600
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(600)
US-09-815-242-6072

Alignment Scores:
Pred. No.: 149
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 10
Gaps: 0

US-10-024-955-7 (1-213) x US-09-815-242-6072 (1-600)

QY 185 ValLeuThrAlaIlePheGln 191
|||||
Db 441 GTTCACGCCCATTTCCAG 421

RESULT 28
US-09-764-853-71/c

; Sequence 71, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PU206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-71

Alignment Scores:
Pred. No.: 149 Length: 600
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-764-853-71 (1-600)

OY 141 AlaleuSerleuGlulIleSer 147
|||||
DB 52 GCCTTATCCCTTGAAATTCT 32

RESULT 29
US-09-770-445-885/c
; Sequence 885, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Kameoka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, MaJa
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 885
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-885

Alignment Scores:
Pred. No.: 189 Length: 777
Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-770-445-885 (1-777)

OY 185 ValLeuThrAlaIlePheGln 191
|||||
DB 445 GTCCTTACCCCATTTCCAA 425

RESULT 30
US-09-799-777-82/c
; Sequence 82, Application US/09799777
; Patent No. US20020091244A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Corley, Neil C.
; Guegler, Karl J.
; Baugh, Mariah
; Sather, Susan
; Shah, Purvi
; TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS
; NUMBER OF SEQUENCES: 154
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/799,777
; FILING DATE: 06-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/002,485
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BILLINGS, LUCY J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0459 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 846 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITUT03
; CLONE: 866885
; SEQUENCE DESCRIPTION: SEQ ID NO: 82 :
US-09-799-777-82

Alignment Scores:
Pred. No.: 204 Length: 846
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-799-777-82 (1-846)

Oy 181 ValLeuSerAspValLeuThr 187
|||||
Db 236 GTGTATCTGATGCTGTGACA 216

RESULT 31
US-09-764-868-29/c
: Sequence 29, Application US/09764868
: Patent No. US20020168711A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT232
: CURRENT APPLICATION NUMBER: US/09/764, 868
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1510
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 29
: LENGTH: 913
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (61)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (758)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (824)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (850)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (858)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (869)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (891)
: OTHER INFORMATION: n equals a,t,g, or c
: OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-29

Alignment Scores:
Pred. No.: 218 Length: 913
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x US-09-764-868-29 (1-913)

Oy 160 ArgGlnPheAlaAsnValVal 166
|||||
Db 222 CGACAGTTTCTGCTAATGTGTA 202

RESULT 32
US-09-815-242-7209/c
: Sequence 7209, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in

: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7209
: LENGTH: 924
: TYPE: DNA
: ORGANISM: Helicobacter pylori
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1) .. (924)
US-09-815-242-7209

Alignment Scores:
Pred. No.: 221 Length: 924
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x US-09-815-242-7209 (1-924)

Oy 179 PheGlyValLeuSerAspVal 185
|||||
Db 630 TTTCGCGTCTGAGTGATGTC 610

RESULT 33
US-09-886-055-290/c
: Sequence 290, Application US/09886055
: Patent No. US20020132273A1
: GENERAL INFORMATION:
: APPLICANT: STREYER, LUBERT
: APPLICANT: ZOZULYA, SERGEY
: TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
: FILE REFERENCE: 078003-0277150
: CURRENT APPLICATION NUMBER: US/09/886,055
: CURRENT FILING DATE: 2001-06-22
: PRIOR APPLICATION NUMBER: 60/213,812
: PRIOR FILING DATE: 2000-06-22
: NUMBER OF SEQ ID NOS: 522
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 290
: LENGTH: 939
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-886-055-290

Alignment Scores:
Pred. No.: 224 Length: 939
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x US-09-886-055-290 (1-939)

OY 106 GIYVALHISASPPILEVAl 112
|
DB 414 GGGGTCATGATGACATGTGA 394

RESULT 34
US-09-778-844-120/c

; Sequence 120, Application US/09778844
; Patent No. US20020150971A1

; GENERAL INFORMATION:

; APPLICANT: JOHANSEN, JEANETTE ELISABETH

; APPLICANT: SCHALLING, MARTIN

; TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES FOR CONTROLLING FOOD

; FILE REFERENCE: 030307/0195

; CURRENT APPLICATION NUMBER: US/09/778, 844

; CURRENT FILING DATE: 2001-02-08

; NUMBER OF SEQ ID NOS: 206

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 120

; LENGTH: 1065

; TYPE: DNA

; ORGANISM: Unknown Organism

; FEATURE:

; OTHER INFORMATION: Description of Unknown Organism: CPLA2-beta, EMBL

US-09-778-844-120

Alignment Scores:

Pred. No.: 251

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 3.29%

DB: 10

Length: 1065

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-024-955-7 (1-213) x US-09-778-844-120 (1-1065)

OY 139 VALVALALALEUSERLEUGlu 145
|
DB 152 GTAGTGGCGCTTCTCTAGAA 132

RESULT 35

US-09-925-299-125/c

; Sequence 125, Application US/09925299

; Patent No. US20020055627A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA102

; CURRENT APPLICATION NUMBER: US/09/925, 299

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124, 270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 125

; LENGTH: 1267

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-299-125

Alignment Scores:

Pred. No.: 293

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 3.29%

DB: 10

Length: 1267

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-024-955-7 (1-213) x US-09-925-299-125 (1-1267)

OY 13 VALALAVALSERAlaSPPro 19
|
DB 1124 GTGGCAGTGTCTGCTGATCCA 1104

RESULT 36

US-09-974-300-5126/c

; Sequence 5126, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.

; APPLICANT: Clausen, Ib Groth

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; FILE REFERENCE: 10085.500-US

; CURRENT APPLICATION NUMBER: US/09/974, 300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680, 598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279, 526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5126

; LENGTH: 1332

; TYPE: DNA

; ORGANISM: Bacillus clausii

US-09-974-300-5126

Alignment Scores:

Pred. No.: 307

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 3.29%

DB: 10

Length: 1332

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-024-955-7 (1-213) x US-09-974-300-5126 (1-1332)

OY 11 ALAPHEVALALAVASERAla 17
|
DB 165 GCTTTCGTTGCGGCTCGGCC 145

RESULT 37

US-09-938-842A-932

; Sequence 932, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Krepes, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE REFERENCE: SCRIPT300-3

; CURRENT APPLICATION NUMBER: US/09/938, 842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227, 866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264, 647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300, 111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 932

; LENGTH: 1479

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-932

Alignment Scores:

Pred. No.: 337

Score: 1479

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 3.29%

DB: 10

Length: 1479

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x US-09-938-842A-932 (1-1479)

Oy 180 GYValleuserasvallen 186

DB 289 GGTGCTCTCAGATCTTCTT 309

RESULT 38

US-09-989-442-64

Sequence 64, Application US/09989442

Publication No. US20030013649A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: P1208

CURRENT APPLICATION NUMBER: US/09/989,442

PRIOR APPLICATION NUMBER: 2001-11-21

PRIOR FILING DATE: 2000-01-31

PRIOR APPLICATION NUMBER: 60/180,628

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: 60/214,886

PRIOR FILING DATE: 2000-06-28

PRIOR APPLICATION NUMBER: 60/217,487

PRIOR FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/225,758

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/220,963

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/217,496

PRIOR FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/225,447

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/218,290

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/225,757

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/226,868

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: 60/216,647

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: 60/225,267

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/216,880

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: 60/225,270

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/251,869

PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/235,834

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: 60/234,274

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: 60/234,223

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: 60/228,924

PRIOR FILING DATE: 2000-08-30

PRIOR APPLICATION NUMBER: 60/224,518

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/236,369

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/224,519

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/220,964

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/241,809

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/249,299

PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836
PRIOR FILING DATE: 2000-09-27

```

; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,065
; PRIOR FILING DATE: 2000-09-14

```

```

Alignment Scores:
Pred. No.: 344
Score: 7.00
Percent Similarity: 100.00%
Length: 1511
Matches: 7
Conservative: 0

```

```

Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 9
Gaps: 0

```

```

US-10-024-955-7 (1-213) x US-09-989-442-64 (1-1511)

```

```

QY 141 AlaleuSerleugluilesr 147
Db 1094 GCCTTATCCCTTGAAATTCT 1114

```

```

RESULT 39

```

```

; US-09-764-853-308
; Sequence 308, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 308
; LENGTH: 1511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1511)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-853-308

```

```

Alignment Scores:

```

```

Pred. No.: 344
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 10
Length: 1511
Matches: 7
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

```

US-10-024-955-7 (1-213) x US-09-764-853-308 (1-1511)

```

```

QY 141 AlaleuSerleugluilesr 147
Db 1094 GCCTTATCCCTTGAAATTCT 1114

```

```

RESULT 40

```

```

; US-09-932-367A-1/c
; Sequence 1, Application US/09932367A
; Publication No. US20030027152A1
; GENERAL INFORMATION:
; APPLICANT: RHODES, Simon J.
; APPLICANT: BRIDWELL, Jeanne L.
; APPLICANT: MEIER, Bradley C.
; APPLICANT: PARKER, Gretchen E.
; APPLICANT: PRICE, Jeffrey R.
; APPLICANT: SHOWALTER, Aaron D.
; APPLICANT: SLOPE, Kyle W.
; TITLE OF INVENTION: GENERATION OF DIAGNOSTIC TOOLS TO ASSAY THE HUMAN
; FILE REFERENCE: LHX3/P-LIM/LIM-3 FACTOR
; CURRENT APPLICATION NUMBER: US/09/932,367A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: PCT/US00/04424
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/121,110
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1605
; TYPE: DNA

```



```
; ORGANISM: Sus scrofa
US-09-932-367A-1

Alignment Scores:
Pred. No.: 363
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 9

US-10-024-955-7 (1-213) x US-09-932-367A-1 (1-1605)

QY 199 ThrlvsvalleuAlaprola 205
DB 643 ACGAAGTACTGCCCCACG 623

RESULT 41
US-09-932-367A-13/c
; Sequence 13, Application US/09932367A
; Publication No. US20030027152A1
; GENERAL INFORMATION:
; APPLICANT: RHODES, Simon J.
; APPLICANT: BRIDWELL, Jeanne L.
; APPLICANT: MEIER, Bradley C.
; APPLICANT: PARKER, Gretchen E.
; APPLICANT: PRICE, Jeffrey R.
; APPLICANT: SHOWALTER, Aaron D.
; APPLICANT: SLOOP, Kyle W.
; TITLE OF INVENTION: GENERATION OF DIAGNOSTIC TOOLS TO ASSAY THE HUMAN
; FILE REFERENCE: 053884-5003
; CURRENT APPLICATION NUMBER: US/09/932,367A
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: PCT/US00/04424
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/121,110
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1658
; TYPE: DNA
; ORGANISM: Sus scrofa
US-09-932-367A-13

Alignment Scores:
Pred. No.: 374
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 9

US-10-024-955-7 (1-213) x US-09-932-367A-13 (1-1658)

QY 199 ThrlvsvalleuAlaprola 205
DB 696 ACGAAGTACTGCCCCACG 676

RESULT 42
US-09-932-367A-15/c
; Sequence 15, Application US/09932367A
; Publication No. US20030027152A1
; GENERAL INFORMATION:
; APPLICANT: RHODES, Simon J.
; APPLICANT: BRIDWELL, Jeanne L.
; APPLICANT: MEIER, Bradley C.
; APPLICANT: PARKER, Gretchen E.
; APPLICANT: PRICE, Jeffrey R.
; APPLICANT: SHOWALTER, Aaron D.
; APPLICANT: SLOOP, Kyle W.
; TITLE OF INVENTION: GENERATION OF DIAGNOSTIC TOOLS TO ASSAY THE HUMAN
```

```
; TITLE OF INVENTION: LHX3/P-LIM/LIM-3 FACTOR
; FILE REFERENCE: 053884-5003
; CURRENT APPLICATION NUMBER: US/09/932,367A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: PCT/US00/04424
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/121,110
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1664
; TYPE: DNA
; ORGANISM: Sus scrofa
US-09-932-367A-15

Alignment Scores:
Pred. No.: 375
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 9

US-10-024-955-7 (1-213) x US-09-932-367A-15 (1-1664)

QY 199 ThrlvsvalleuAlaprola 205
DB 702 ACGAAGTACTGCCCCACG 682

RESULT 43
US-10-116-252-2/c
; Sequence 2, Application US/10116252
; Publication No. US20030028008A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Seven Transmembrane Receptor Polynucleotides,
; FILE REFERENCE: PT007P1
; CURRENT APPLICATION NUMBER: US/10/116,252
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US/09/711,909
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: PCT/US00/13737
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 60/135,167
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 60/143,616
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 60/152,934
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: 60/189,029
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-116-252-2

Alignment Scores:
Pred. No.: 399
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.29%
DB: 9

US-10-024-955-7 (1-213) x US-10-116-252-2 (1-1782)

QY 169 lIeglyglyleuserlIeueu 175
lIeglyglyleuserlIeueu 175
```

DB 1624 ATTGGTGTCTCTATATTG 1604

RESULT 44

US-09-938-842A-2865/C
; Sequence 2865, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2865
; LENGTH: 1902
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2865

Alignment Scores:

Pred. No.: 423 Length: 1902
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x US-09-938-842A-2865 (1-1902)

QY 181 ValLeuseraPyValLeuthr 187
|||||

DB 488 GTTTGAGTGACGTGTGACT 468

RESULT 45

US-09-733-757-1/C
; Sequence 1, Application US/09733757
; Patent No. US20020102544A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS AND M
; FILE REFERENCE: A-69796/DOB/JJD
; CURRENT APPLICATION NUMBER: US/09/733,757
; CURRENT FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2103
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70)..(2103)
US-09-733-757-1

Alignment Scores:

Pred. No.: 463 Length: 2103
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0

DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-733-757-1 (1-2103)

QY 160 ArgGlnPheAlaAsnValVal 166
|||||

DB 1047 CGACAGTTGCTATGTGGTA 1027

RESULT 46

US-09-742-869-1/C
; Sequence 1, Application US/09742869
; Patent No. US20010016319A1
; GENERAL INFORMATION:
; APPLICANT: SHABON, USMAN
; APPLICANT: BERGSMAN, DERK J.
; APPLICANT: SATHE, GANESH M.
; TITLE OF INVENTION: CLONING OF A NOVEL G-PROTEIN COUPLED 7TM
; FILE REFERENCE: ATG-50047-DIC1
; CURRENT APPLICATION NUMBER: US/09/742,869
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 08/788,539
; PRIOR FILING DATE: 1997-01-24
; PRIOR APPLICATION NUMBER: 09/168,417
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2273
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-742-869-1

Alignment Scores:
Pred. No.: 497 Length: 2273
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-742-869-1 (1-2273)

QY 141 AlaLeuserLeugLuiser 147
|||||

DB 966 GCCTTATCCTGGAATTCT 946

RESULT 47

US-09-964-238-3/C
; Sequence 3, Application US/09964238
; Patent No. US20020035246A1
; GENERAL INFORMATION:
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie
; TITLE OF INVENTION: SMAD-INTERACTING POLYPEPTIDES AND THEIR USE
; FILE REFERENCE: 2676-423205
; CURRENT APPLICATION NUMBER: US/09/964,238
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/449,285
; PRIOR FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: PCT/EP98/03193
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 97201645.5
; PRIOR FILING DATE: 1997-06-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 2959
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-964-238-3

Alignment Scores:

Pred. No.: 630 Length: 2959

Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-964-238-3 (1-2959)

OY 8 AlaAlaValAlaPheValAla 14
Db 1823 GCACGCGTAGCCTTTGTTGCG 1803

RESULT 48
US-09-879-248-7/c
Sequence 7, Application US/09879248
Patent No. US20020062500A1
GENERAL INFORMATION:
APPLICANT: Fan, Hao
APPLICANT: Wei, Zhong-Min
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
FILE REFERENCE: 21829/81
CURRENT APPLICATION NUMBER: US/09/879, 248
PRIOR FILING DATE: 2001-06-12
PRIOR APPLICATION NUMBER: 60/212, 211
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 7
LENGTH: 5517
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-879-248-7

Alignment Scores:
Pred. No.: 1.1e+03 Length: 5517
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x US-09-879-248-7 (1-5517)

OY 138 PheValValAlaIleuSerLeu 144
Db 4778 TTCGTTGGCGCTATTCCTTA 4758

RESULT 49
US-10-114-170-102/c
Sequence 102, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: NO. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114.170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 7886
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-114-170-102

Alignment Scores:
Pred. No.: 1.52e+03 Length: 7886
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x US-10-114-170-102 (1-7886)

OY 4 PheLeuLeuIleAlaVal 10
Db 5294 TTTTGCATGATTCGCGCGTG 5274

RESULT 50
US-10-033-190-3/c
Sequence 3, Application US/10033190
Patent No. US20020133848A1
GENERAL INFORMATION:
APPLICANT: Exelixis Plant Sciences, Inc.
TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF AN ANTHOCYANIN MUTANT
FILE REFERENCE: EP01-002C
CURRENT APPLICATION NUMBER: US/10/033,190
PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: US 60/244,685
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 10078
TYPE: DNA
ORGANISM: PAG3202
US-10-033-190-3

Alignment Scores:
Pred. No.: 1.9e+03 Length: 10078
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.29% Indels: 0
DB: 12 Gaps: 0

US-10-024-955-7 (1-213) x US-10-033-190-3 (1-10078)

OY 179 PheGlyValLeuSerAspVal 185
Db 5207 TTCGGCGTGCCTCCGATGTC 5187

Search completed: February 20, 2003, 20:13:20
Job time : 130 secs

• • • •

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 18:32:25; Search time 1471 Seconds

(without alignments)
2345.098 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213

Sequence: 1 MMKFLIAAVAFVAVSADPL.....VRKEMTKVLAPAFKRELEKN 213

Scoring table:

OLIGO	
Xgapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n model -DEV=x1p
-Q=/cgn2_1/USPTO.spool/US10024955/runatc_14022003.130736_24263/app-query.fasta.1.391
-DB=EST -QFMT=fastlap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10024955.8CGN.1.1.2874.6runatc_14022003.130736_24263 -NCPU=6 -ICPU=3
-NO_ALPYX -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7
```

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlu:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	9	4.2	530	17 BH214589	BH214589 KG00824 D
2	9	4.2	635	10 AV705987	AV705987 AV705987
3	8	3.8	114	17 BH196069	BH196069 TC3-1.71J20
4	8	3.8	137	13 BJ362558	BJ362558 BJ362558
5	8	3.8	151	13 BJ387310	BJ387310 BJ387310
6	8	3.8	172	9 A1756312	A1756312 ETESTE42
7	8	3.8	217	13 BJ358854	BJ358854 BJ358854
8	8	3.8	218	13 BJ387664	BJ387664 BJ387664
9	8	3.8	227	13 BJ364168	BJ364168 BJ364168
10	8	3.8	229	10 BE076526	BE076526 CM1-B7059
11	8	3.8	240	9 A0072751	A0072751 A0072751
12	8	3.8	241	12 BF606164	BF606164 273125 MA
13	8	3.8	254	13 BJ388702	BJ388702 BJ388702
14	8	3.8	257	13 BJ389545	BJ389545 BJ389545
15	8	3.8	269	10 AV621032	AV621032 AV621032
16	8	3.8	285	13 BM288495	BM288495 529776 MA
17	8	3.8	292	17 A0076852	A0076852 CIT-HSP-2
18	8	3.8	330	13 BJ362907	BJ362907 BJ362907
19	8	3.8	341	12 BF361558	BF361558 poc-1.58-
20	8	3.8	341	13 BJ358546	BJ358546 BJ358546
21	8	3.8	341	13 BJ386928	BJ386928 BJ386928
22	8	3.8	349	13 BJ387265	BJ387265 BJ387265
23	8	3.8	352	13 BJ325557	BJ325557 BJ325557
24	8	3.8	361	9 A0062214	A0062214 A0062214
25	8	3.8	365	13 BJ326257	BJ326257 BJ326257
26	8	3.8	366	17 A0191207	A0191207 HS-2195_B
27	8	3.8	367	17 A0323927	A0323927 RPT11-10
28	8	3.8	368	9 A1065613	A1065613 ag90e10.x
29	8	3.8	378	13 BJ359974	BJ359974 BJ359974
30	8	3.8	394	13 BJ326200	BJ326200 BJ326200
31	8	3.8	400	10 BB734318	BB734318 BB734318
32	8	3.8	400	13 BJ388026	BJ388026 BJ388026
33	8	3.8	403	13 BJ329427	BJ329427 BJ329427
34	8	3.8	406	13 BM531053	BM531053 fw2a07.Y
35	8	3.8	413	13 BJ391088	BJ391088 BJ391088
36	8	3.8	415	13 BI200525	BI200525 n3e10fs.r
37	8	3.8	415	13 BI418268	BI418268 L1NEST46h
38	8	3.8	416	13 BJ391085	BJ391085 BJ391085
39	8	3.8	419	14 W32146	W32146 2b97906.r1
40	8	3.8	425	13 BJ388944	BJ388944 BJ388944
41	8	3.8	438	13 BJ386722	BJ386722 BJ386722
42	8	3.8	441	13 BJ388417	BJ388417 BJ388417
43	8	3.8	442	13 BJ358769	BJ358769 BJ358769
44	8	3.8	443	14 BQ146887	BQ146887 NF029D03F
45	8	3.8	445	13 BJ358897	BJ358897 BJ358897

ALIGNMENTS

```
RESULT 1
LOCUS BH214589
DEFINITION BH214589 530 bp DNA linear GSS 02-NOV-2001
lines Drosophila melanogaster P(SUPOR-P) P element insertion
5' and 3' ends of P element, DNA sequence.
ACCESSION BH214589
VERSION BH214589.1 GI:16596872
KEYWORDS GSS.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 530)
```

AUTHORS
Levis,R., Hoskins,R., Liao,G., Mozen,N., Tsang,G., He,Y., Karpen
,G., Belien,H., Rubin,G. and Spradling,A.
The Berkeley Drosophila Genome Project Gene Disruption Project
JOURNAL
Unpublished (2001)
COMMENT
Contact: Gerald Rubin
Berkeley Drosophila Genome Project
University of California, Berkeley
LSA Building, Berkeley, CA 94720-3200, USA
Fax: 5106439947
Email: gerry@fruitfly.berkeley.edu
Sequence recovery method was inverse PCR.
Sequence orientation is forward strand relative to 5' end of P
element
The P element insertion position is base 103 in the 530 bases.
This insertion position refers to the first base of the 8 base
target recognition sequence.
Class: transposon-tagged.
FEATURES
source
1..530
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="Drosophila melanogaster P(SUPor-P) P element
insertion lines"
/note="Inverse PCR was performed on Drosophila
melanogaster strains each of which contains one or more
P(SUPor-P) P-element transposon insertion. The resultant
fragment for each strain was directly sequenced to
determine the genomic sequence at the site of insertion.
Details of the protocols used can be found at
http://www.fruitfly.org/about/methods/inverse.pcr.html."

BASE COUNT 127 a 122 c 127 g 154 t
ORIGIN

Alignment Scores:
Pred. No.: 54.2 Length: 530
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 17 Gaps: 0

US-10-024-955-7 (1-213) x BH214589 (1-530)
Qy 8 A1a1aVala1aPheVala1aVala1aSer 16
Db 26 GCAGCAGTCGCATTTGCGCTGTCTC 52

RESULT 2
AV705987/c 635 bp mRNA linear EST 09-OCT-2000
LOCUS AV705987 ADB Homo sapiens cDNA ADBD109 5', mRNA sequence.
ACCESSION AV705987
VERSION AV705987.1 GI:10723276
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 635)
Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao
,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA ADB clones
unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chnc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
Location/Qualifiers
1..635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ADB109"
/clone_lib="ADB"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 198 a 133 c 98 g 204 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 68.1 Length: 635
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.23% Indels: 0
DB: 10 Gaps: 0

US-10-024-955-7 (1-213) x AV705987 (1-635)
Qy 141 A1a1eSer1eUc1u1eSer1aSp1u 149
Db 502 GCCTTTCTCTCGAGATCTCTGATGAA 476

RESULT 3
BH196069 114 bp DNA linear GSS 24-OCT-2001
LOCUS BH196069
DEFINITION TC3-71J20.TP TC3 Trypanosoma cruzi genomic clone TC3-71J20, DNA
sequence.
ACCESSION BH196069
VERSION BH196069.1 GI:16364090
KEYWORDS GSS.
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 114)
REFERENCE Kluge,S., Edwards,K.E., Nilsson,D., Bontempi,E.J., Myler,P., Stuart
AUTHORS K., Ghedin,E., El-Sayed,N.M. and Andersson,B.
TITLE Clustering and analysis of BAC-end and GSS sequences from
Trypanosoma cruzi
JOURNAL Unpublished (2001)
COMMENT Other GSSs: TC3-71J20.TV
Contact: Bjorn Andersson
Department of Genetics and Pathology
Uppsala University
Rudbeck Laboratory, SE-751 85, Uppsala, Sweden
Tel: 46 18 471 4107
Fax: 46 18 471 4808
Email: bjorn.andersson@genpat.uu.se
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: SP6
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..114
/organism="Trypanosoma cruzi"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-71J20"
/clone_lib="TC3"
/note="Vector: pBelobAC11; Site_1: Hin dIII; Constructed
for Uppsala University by Marie-Christine Le Paslier in
the laboratory of Denis Le Paslier at the Centre d'Etude
du Polymorphisme Humain (CEPH), Paris, France. Briefly,
Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
from Dr. Franco da Silveira) was partially digested with
Hin dIII. High molecular weight fragments were ligated in

pbe10BAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 x the haploid genome."

BASE COUNT 19 a 23 c 36 g 36 t

ALIGNMENT SCORES:

Pred. No.: 83 Length: 114
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 17 Gaps: 0

US-10-024-955-7 (1-213) x BH196069 (1-114)

OY 180 GYVAILEuserAspVallleuthr 187

DB 64 GGAGTACTTTCGACGTACTTACT 87

RESULT 4

LOCUS BJ362558 137 bp mRNA linear EST 07-MAR-2002
DEFINITION BJ362558 Dictyostelium discoideum cDNA library, CF Dictyostelium
discoideum cDNA clone ddc22m10 5', mRNA sequence.

ACCESSION BJ362558

VERSION BJ362558.1 GI:19262153

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 137)
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
stage

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

FEATURES

source

1..137
/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="ddc22m10"

/sex="mat A"

/dev_stage="Culmination stage"

BASE COUNT 41 a 22 c 17 g 55 t 2 others

ORIGIN

ALIGNMENT SCORES:

Pred. No.: 105 Length: 137
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ362558 (1-137)

OY 123 GYAspleuHISProThrThrHis 130

DB 110 GGTGATTTCGACGTACTTACT 133

RESULT 5

LOCUS BJ387310 151 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ387310 Dictyostelium discoideum cDNA library, SF Dictyostelium

discoideum cDNA clone dds1e17 5', mRNA sequence.

ACCESSION BJ387310

VERSION BJ387310.1 GI:19296694

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 151)
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

TITLE Full length cDNA of Dictyostelium discoideum at the slug stage

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

FEATURES

source

1..151
/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="dds1e17"

/clone_lib="Dictyostelium discoideum cDNA library, SF"

/sex="mat A"

/dev_stage="Slug stage"

BASE COUNT 44 a 25 c 22 g 59 t 1 others

ORIGIN

ALIGNMENT SCORES:

Pred. No.: 119 Length: 151
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ387310 (1-151)

OY 123 GYAspleuHISProThrThrHis 130

DB 91 GGTGATTTCGACGTACTTACT 114

RESULT 6

AT1756312/c

LOCUS AT1756312 172 bp mRNA linear EST 18-JAN-2000

DEFINITION EtkSTEa42d05.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA

5', mRNA sequence.

ACCESSION AT1756312

VERSION AT1756312.1 GI:5150035

KEYWORDS EST.

SOURCE Eimeria tenella.

ORGANISM Eimeria tenella.

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

Eimeria

1 (bases 1 to 172)

Liberator, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,

Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen

, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey

, N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson

, Y., Cardenas, M., McCann, R., Waterson, R., Wilson, R. and Sibley, D.

WashU-Merck Eimeria tenella project

unpublished (1999)

CONTACT: David Sibley, Ph.D.

WashU-Merck Eimeria tenella project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Contact David Sibley (toxosteborcin.wustl.edu) for further

information relating to organism, libraries, or clone availability.

Seq primer: -40RP from Gibco
High quality sequence stop: 161.
Location/Qualifiers
1. .172

/organism="Elmeria tenella"
/strain="LS18"

/db_xref="taxon:5802"
/clone_lib="Elmeria S5-2 Sporozoite stage"

/dev_stage="Sporozoite"

/lab_host="SOLR E. coli"

/note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of E. tenella grown in chickens. cDNA
was synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,
EcoRI adapters were ligated to the cDNA and products were
size-selected on Sephacryl S500. cDNAs were digested with
EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
Clones were converted to phagemids by mass excision using
Exassist helper phage and SOLR cells (Stratagene).
Insert sizes range from 1.2-2.9 kb."

BASE COUNT 67 a 42 c 45 g 18 t
ORIGIN

Alignment Scores:

Pred. No.: 140 Length: 172
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x AT756312 (1-172)

QY 9 AlavAlaIapheValAlaValSer 16

DB 31 GCTGTTGCTTGTGCTGTTCC 8

RESULT 7

LOCUS BJ358854 217 bp mRNA linear EST 07-MAR-2002
DEFINITION BJ358854 Dictyostelium discoideum cDNA library, CF Dictyostelium

ACCESSION BJ358854 Dictyostelium cDNA ddc1113 5', mRNA sequence.

VERSION BJ358854.1 GI:19258449

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum; Dictyostelid; Dictyostelium.

REFERENCE 1 (bases 1 to 217)

AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.

TITLE Full length cDNA of Dictyostelium discoideum at the culmination

JOURNAL stage

COMMENT Unpublished (2002)

CONTACT: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6855

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

Location/Qualifiers

1. .217

/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone_lib="ddc1113"

/clone_lib="Dictyostelium discoideum cDNA library, CF"

/sex="mat A"

/dev_stage="Culmination stage"

BASE COUNT 69 a 30 c 23 g 95 t

ORIGIN

Alignment Scores:

Pred. No.: 188 Length: 217
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ358854 (1-217)

QY 123 GLYAspleuHisProthrTrHis 130

DB 162 GGTGATTGCATCCACTACTCAT 185

RESULT 8

LOCUS BJ387664 218 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ387664 Dictyostelium discoideum cDNA library, SF Dictyostelium

ACCESSION BJ387664 Dictyostelium cDNA dds4101 5', mRNA sequence.

VERSION BJ387664.1 GI:19297048

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum; Dictyostelid; Dictyostelium.

REFERENCE 1 (bases 1 to 218)

AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.

TITLE Full length cDNA of Dictyostelium discoideum at the slug stage

JOURNAL Unpublished (2002)

CONTACT: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6855

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

Location/Qualifiers

1. .218

/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone_lib="dds4101"

/clone_lib="Dictyostelium discoideum cDNA library, SF"

/sex="mat A"

/dev_stage="Slug stage"

BASE COUNT 66 a 35 c 38 g 78 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 189 Length: 218

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.76% Indels: 0

DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ387664 (1-218)

QY 123 GLYAspleuHisProthrTrHis 130

DB 112 GGTGATTGCATCCACTACTCAT 135

RESULT 9

LOCUS BJ364168 227 bp mRNA linear EST 08-MAR-2002

DEFINITION BJ364168 Dictyostelium discoideum cDNA library, CF Dictyostelium

ACCESSION BJ364168 Dictyostelium cDNA clone ddc30110 5', mRNA sequence.

VERSION BJ364168.1 GI:19273468

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum; Dictyostelid; Dictyostelium.

REFERENCE 1 (bases 1 to 227)
 AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
 TITLE Full length cDNA of Dictyostellium discoideum at the culmination stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 FEATURES
 source
 1..227
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone_lib="ddc30j10"
 /clone_lib="Dictyostellium discoideum cDNA library, CF"
 /sex="mat A"
 /dev_stage="Culmination stage"
 BASE COUNT 55 a 47 c 44 g 72 t 9 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 199 Length: 227
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 13 Gaps: 0
 US-10-024-955-7 (1-213) x BJ364168 (1-227)
 QY 123 GIYASPLEUHISPR0THRHIS 130
 DB 66 GGTGATTTCATCCATCCTCAT 89
 RESULT 10
 BE076526/c
 LOCUS BE076526 229 bp mRNA linear EST 09-JUN-2000
 DEFINITION CM1-BT0596-110100-089-d03 BT0596 Homo sapiens CDNA, mRNA sequence.
 ACCESSION BE076526
 VERSION BE076526.1 GI:8426435
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 229)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SF,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-cm1-BT0596-110
 100-089-d03&t3=2000-01-11&t4=1)
 Seq primer: puc 18 forward

FEATURES
 source
 1..229
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0596"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site:1: SmaI; Site:2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 ,716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 BASE COUNT 74 a 52 c 48 g 75 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 201 Length: 229
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 10 Gaps: 0
 US-10-024-955-7 (1-213) x BE076526 (1-229)
 QY 12 PheValAlaValSerAlaAspPro 19
 DB 153 TTTGTGCTGTGTGTCAGTCATCCG 130
 RESULT 11
 AU072751
 LOCUS AU072751 240 bp mRNA linear EST 24-JUN-1999
 DEFINITION AU072751 Dictyostellium discoideum SS (H.Urushihara) Dictyostellium
 discoideum cDNA clone SSA753, mRNA sequence.
 ACCESSION AU072751
 VERSION AU072751.1 GI:5179172
 KEYWORDS EST.
 SOURCE Dictyostellium discoideum.
 ORGANISM Dictyostellium discoideum.
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 REFERENCE 1 (bases 1 to 240)
 AUTHORS Urushihara,H.
 TITLE Developmental cDNA in Dictyostellium discoideum (1999)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Hideo Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
 Tel: 81-298-53-4664
 Fax: 81-298-53-6614
 Email: hideo@biol.tsukuba.ac.jp
 PROJECT Dictyostellium discoideum cDNA project in Japan.
 FEATURES
 source
 1..240
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone_lib="SSA753"
 /clone_lib="Dictyostellium discoideum SS (H.Urushihara)"
 /dev_stage="slug"
 BASE COUNT 72 a 45 c 48 g 75 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 213 Length: 240
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x AU072751 (1-240)

QY 123 GLYAspleuHisProThrHis 130
 DB 2 GGTGATTGCATCCACTACTCAT 25

RESULT 12
 BE606164/c 241 bp mRNA linear EST 25-APR-2001
 LOCUS
 DEFINITION B273125 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BF606164
 VERSION BF606164.1 GI:11706495
 KEYWORDS
 SOURCE
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 241)

REFERENCE
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Choj,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Perlea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
 TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
 JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE
 COMMENT

BASE COUNT
 ORIGIN
 FEATURES
 source
 1.241
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone.lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site_1: NotI; Site_2: SalI; library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."
 Seq primer: ATTAGTGCACATATAG.
 Location/Qualifiers

BASE COUNT 71 a 39 c 42 g 89 t
 ORIGIN

Alignment Scores:
 Pred. No.: 214 Length: 241
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 12 Gaps: 0

US-10-024-955-7 (1-213) x BF606164 (1-241)

QY 58 PheGLuArgHisValGlyIleVal 65
 DB 63 TTTGAAGCATGTGGGATTGTT 40

RESULT 13
 BJ388702

LOCUS BJ388702 254 bp mRNA linear EST 08-MAR-2002
 DEFINITION BJ388702 Dictyostelium discoideum cDNA library, SF Dictyostelium
 accession cDNA clone dds6120 5', mRNA sequence.
 ACCESSION BJ388702
 VERSION BJ388702.1 GI:19299788
 KEYWORDS
 SOURCE
 ORGANISM Dictyostelium discoideum.
 Dictyostelium discoideum.
 Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.

REFERENCE
 AUTHORS 1 (bases 1 to 254)
 TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 source
 1.254
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone.lib="dds6120"
 /clone.lib="Dictyostelium discoideum cDNA library, SF"
 /sex="mat A"
 /dev_stage="slug stage"

BASE COUNT 77 a 44 c 37 g 95 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 229 Length: 254
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ388702 (1-254)

QY 123 GLYAspleuHisProThrHis 130
 DB 135 GGTGATTGCATCCACTACTCAT 158

RESULT 14
 BJ389545

LOCUS BJ389545 257 bp mRNA linear EST 08-MAR-2002
 DEFINITION BJ389545 Dictyostelium discoideum cDNA library, SF Dictyostelium
 accession cDNA clone dds19m01 5', mRNA sequence.
 ACCESSION BJ389545
 VERSION BJ389545.1 GI:19300631
 KEYWORDS
 SOURCE
 ORGANISM Dictyostelium discoideum.
 Dictyostelium discoideum.
 Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.

REFERENCE
 AUTHORS 1 (bases 1 to 257)
 TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 source
 1.257
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"

```

/clone="dcds19m01"
/clone_lib="Dictyostellium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"
BASE COUNT      80 a      38 c      34 g      105 t
ORIGIN

Alignment Scores:
Pred. No.:      233      Length:      257
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      3.76%      Indels:      0
DB:              13      Gaps:      0

US-10-024-955-7 (1-213) x BJ389545 (1-257)

Oy      123 GtAspLeuHsPrcHrThrHis 130
        |||||||
Db      162 GGTGATTTCATGCCACTACTCAT 185

RESULT 15
LOCUS      AV621032      269 bp      mRNA      linear      EST 15-DEC-2000
DEFINITION      AV621032 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
                reinhardtii cDNA clone LC030b12_r 5', mRNA sequence.
ACCESSION      AV621032
VERSION      AV621032.1 GI:10770207
KEYWORDS      EST.
SOURCE      Chlamydomonas reinhardtii.
ORGANISM      Chlamydomonas reinhardtii
                Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
                Chlamydomonadaceae; Chlamydomonas.
REFERENCE      1 (bases 1 to 269)
AUTHORS      Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
                Nakamura,Y. and Tabata,S.
TITLE      Generation of expressed sequence tags from low-CO2 and high-CO2
                adapted cells of Chlamydomonas reinhardtii
JOURNAL      DNA Res. 7 (5), 305-307 (2000)
MEDLINE      20539644
COMMENT      Contact: Erika Asamizu
                The First Laboratory for Plant Gene Research
                Kazusa DNA Research Institute
                Yana 153-2-3, Kisarazu, Chiba 292-0812, Japan
                Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
    source
        1..269
            /organism="Chlamydomonas reinhardtii"
            /strain="C9"
            /db_xref="taxon:3055"
            /clone="LC030b12_r"
            /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI; The cDNA library was constructed from cells cultured
            in a carbon stress acclimatized condition in which carbon
            dioxide concentration in the bubbling gas was changed from
            5% to 0.04%"
BASE COUNT      53 a      85 c      84 g      47 t
ORIGIN

Alignment Scores:
Pred. No.:      247      Length:      269
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      3.76%      Indels:      0
DB:              10      Gaps:      0

US-10-024-955-7 (1-213) x AV621032 (1-269)

Oy      5 LeuLeuIleAlaIleAlaIleAlaIle 12
        |||||||
Db      155 CTTCATTCATTCGCGCGTTCCTTC 132

```

```

RESULT 16
LOCUS      BM288495/c      285 bp      mRNA      linear      EST 28-DEC-2001
DEFINITION      BM288495 529776 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      BM288495
VERSION      BM288495.1 GI:17997521
KEYWORDS      EST.
SOURCE      COW.
ORGANISM      Bos taurus
                Eukaryota: Metazoa: Chordata: Cranialia: Vertebrata: Euteleostomi:
                Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
                Bovidae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 285)
AUTHORS      Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
                Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
                G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
                Pettea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
                Keefe,J.W.
TITLE      Sequence evaluation of four pooled-tissue normalized bovine cDNA
                libraries and construction of a gene index for cattle
JOURNAL      Genome Res. 11 (4), 626-630 (2001)
MEDLINE      21180013
COMMENT      Contact: Smith TPL
                USDA, ARS, US Meat Animal Research Center
                PO Box 166, Clay Center, NE 68933-0166, USA
                Tel: 402 762 4366
                Fax: 402 762 4390
                Email: smith@email.marc.usda.gov
                Single pass sequencing. Bases called and alt. trimmed with phred
                v0.980904.e. Vector identified by cross_match with the -minscore 18
                and -mismatch 12 options.
                PCR primers
                FORWARD: AGCAACACGCTATGACCAT
                BACKWARD: GTTTCACGTCACGACG
                Plate: 141 row: G column: 17
                Seq primer: ATTAGGTGACACTATAG.
FEATURES
    source
        1..285
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 3BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            Library made from pooled tissue from marrow, alveolar
            macrophage, ovary, fetal semitendinosus muscle, and fetal
            longissimus muscle."
BASE COUNT      82 a      54 c      51 g      98 t
ORIGIN

Alignment Scores:
Pred. No.:      265      Length:      285
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      3.76%      Indels:      0
DB:              13      Gaps:      0

US-10-024-955-7 (1-213) x BM288495 (1-285)

Oy      58 PheGluArgHisValGlyIleVal 65
        |||||||
Db      109 TTGAAAGCAGTGGGATGTGT 86

RESULT 17
LOCUS      A0076852      292 bp      DNA      linear      GSS 20-AUG-1998
DEFINITION      CIT-HSP-2361M9.TR CIT-HSP Homo sapiens genomic clone 2361M9, DNA
                sequence.
ACCESSION      A0076852
VERSION      A0076852.1 GI:3438036
KEYWORDS      GSS.

```

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 292)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Map Building
Use of a random human BAC End Sequence Database for Sequence-Ready
Unpublished (1998)
JOURNAL COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="2361M9"
/clone_lib="CIR-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelovAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 91 a 54 c 43 g 104 t
ORIGIN

Alignment Scores:
Pred. No.: 274 Length: 292
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 17 Gaps: 0

US-10-024-955-7 (1-213) x AQ076852 (1-292)

QY 168 HisIIcGlycIeuSerIleu 175
|||||
Db 181 CATATTGAGGCTCATCTACTT 204

RESULT 18
BJ362907 330 bp mRNA linear EST 07-MAR-2002
BJ362907 Dictyostelium discoideum cDNA library, CF Dictyostelium
discoideum cDNA clone ddc24b03 5', mRNA sequence.
ACCESSION BJ362907
VERSION BJ362907.1 GI:19262502
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 330)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shio-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
stage
Unpublished (2002)
JOURNAL COMMENT Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES
source Location/Qualifiers
1..330
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc24b03"
/clone_lib="Dictyostelium discoideum cDNA library, CF"
/sex="mat A"
/dev_stage="Culmination stage"

BASE COUNT 98 a 61 c 52 g 119 t
ORIGIN

Alignment Scores:
Pred. No.: 319 Length: 330
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ362907 (1-330)

QY 123 GYAspleuHisProThrHis 130
|||||
Db 151 GGTGATTGCATCCACTACTCAT 174

RESULT 19
BF361558 341 bp mRNA linear EST 24-NOV-2000
BF361558
LOCUS pot-1-58-1 Differentially expressed cDNA library of optic tectum of
DEFINITION pigeon with removal of left retina Columbia livia cDNA, mRNA
sequence.
ACCESSION BF361558
VERSION BF361558.1 GI:11323575
KEYWORDS EST.
SOURCE domestic pigeon.
ORGANISM Columbia livia
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
REFERENCE 1 (bases 1 to 341)
AUTHORS Lu,L., Xu,L., Yang,C., Li,X. and Wang,Y.
TITLE A profile of differentially expressed genes in p10 pigeon optic
tectum with removal of left retina after posthatching first day by
SSH

JOURNAL COMMENT Unpublished (2000)
CONTACT: Lixia Lu
Department of Biochemistry
Shanghai Fiedao University Medical School
Gonghexin Rd.1238, Shanghai, 200070, China
Email: jlnroxiad@public.sta.net.cn.

FEATURES
source Location/Qualifiers
1..341
/organism="Columbia livia"
/db_xref="taxon:8932"
/clone_lib="Differentially expressed cDNA library of optic
tectum of pigeon with removal of left retina"
/sex="female and male"
/tissue_type="optic tectum"
/dev_stage="10 day postnatal"
/note="Differentially expressed cDNA library of optic
tectum of pigeon with removal of left retina"

BASE COUNT 88 a 74 c 58 g 121 t
ORIGIN

Alignment Scores:
Pred. No.: 333 Length: 341
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 12 Gaps: 0

US-10-024-955-7 (1-213) x BF361558 (1-341)

QY 172 Leuser1leuaspProtlepHe 179
|||||
Db 146 TTGAGTATTGATCCACTACTATTT 169

RESULT 20
BJ38546 341 bp mRNA linear EST 07-MAR-2002
LOCUS BJ38546 Dictyostelium discoideum cDNA library, CF Dictyostelium
DEFINITION BJ38546 Dictyostelium discoideum cDNA clone ddc10g04 5', mRNA sequence.
ACCESSION BJ38546
VERSION BJ38546.1 GI:19258141
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
REFERENCE 1 (bases 1 to 341)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1. 341
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc10g04"
/clone_lib="Dictyostelium discoideum cDNA library, CF"
/sex="mat A"
/dev_stage="Culmination stage"

BASE COUNT 105 a 56 c 47 g 131 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 333 Length: 341
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x BJ38546 (1-341)

QY 123 GlyaspLeuHISProTHrHis 130
|||||
Db 162 GGTGATTGATCCACTACTCAT 185

RESULT 21
BJ386928 341 bp mRNA linear EST 08-MAR-2002
LOCUS BJ386928 Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION BJ386928 Dictyostelium discoideum cDNA clone dds1n07 5', mRNA sequence.
ACCESSION BJ386928
VERSION BJ386928.1 GI:19296312
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
REFERENCE 1 (bases 1 to 341)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan

FEATURES
SOURCE
1. 341
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc10g04"
/clone_lib="Dictyostelium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"

BASE COUNT 105 a 57 c 47 g 130 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 333 Length: 341
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x BJ386928 (1-341)

QY 123 GlyaspLeuHISProTHrHis 130
|||||
Db 162 GGTGATTGATCCACTACTCAT 185

RESULT 22
BJ387265 349 bp mRNA linear EST 08-MAR-2002
LOCUS BJ387265 Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION BJ387265 Dictyostelium discoideum cDNA clone dds1n04 5', mRNA sequence.
ACCESSION BJ387265
VERSION BJ387265.1 GI:19296649
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
REFERENCE 1 (bases 1 to 349)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1. 349
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds1n04"
/clone_lib="Dictyostelium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"

BASE COUNT 106 a 61 c 50 g 131 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 343 Length: 349
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
Gaps: 0

US-10-024-955-7 (1-213) x BJ387265 (1-349)

OY 123 G1YAsPLeuH1sProThrThrH1s 130
 |||||||||||||||||||
 DB 171 GGTGATTTCATCCACCTACTCAT 194

RESULT 23
 BJ325557

LOCUS BJ325557 352 bp mRNA linear EST 05-MAR-2002

DEFINITION BJ325557 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda1k09 5', mRNA sequence.

ACCESSION BJ325557

VERSION BJ325557.1 GI:19155687

KEYWORDS EST

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum

REFERENCE 1 (bases 1 to 352)
 Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
 Full length cDNA of Dictyostelium discoideum at the aggregation stage

AUTHORS Unpublished (2002)

TITLE Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 source 1..352
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone_lib="ddal1k09"
 /clone_lib="Dictyostelium discoideum cDNA library, AF"
 /sex="mat A"
 /dev_stage="Aggregation stage"

BASE COUNT 106 a 58 c 52 g 135 t 1 others

ORIGIN

Alignment Scores:
 Pred. No.: 347 Length: 352
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ325557 (1-352)

OY 123 G1YAsPLeuH1sProThrThrH1s 130
 |||||||||||||||||||
 DB 180 GGTGATTTCATCCACCTACTCAT 203

RESULT 24
 AU062214

LOCUS AU062214 361 bp mRNA linear EST 20-MAY-1999

DEFINITION AU062214 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium discoideum cDNA clone SLI392, mRNA sequence.

ACCESSION AU062214

VERSION AU062214.1 GI:4883318

KEYWORDS EST

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum

REFERENCE 1 (bases 1 to 361)
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitera,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochial,i.H. and Tanaka,Y.
 Developmental cDNA in Dictyostelium discoideum
 Unpublished (1998)
 Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
 Tel: 81-298-53-4664
 Fax: 81-298-53-6614
 Email: hideko@biol.tsukuba.ac.jp
 PROJECT - Dictyostelium discoideum cDNA project in Japan.

FEATURES
 source Location/Qualifiers
 1..361
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SLI392"
 /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
 /dev_stage="slug"

BASE COUNT 113 a 61 c 59 g 128 t

ORIGIN

Alignment Scores:
 Pred. No.: 358 Length: 361
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x AU062214 (1-361)

OY 123 G1YAsPLeuH1sProThrThrH1s 130
 |||||||||||||||||||
 DB 128 GGTGATTTCATCCACCTACTCAT 151

RESULT 25
 BJ326257

LOCUS BJ326257 365 bp mRNA linear EST 05-MAR-2002

DEFINITION BJ326257 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda15d03 5', mRNA sequence.

ACCESSION BJ326257

VERSION BJ326257.1 GI:19156387

KEYWORDS EST

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum

REFERENCE 1 (bases 1 to 365)
 Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
 Full length cDNA of Dictyostelium discoideum at the aggregation stage

AUTHORS Unpublished (2002)

TITLE Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 source Location/Qualifiers
 1..365
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dda15d03"
 /clone_lib="Dictyostelium discoideum cDNA library, AF"
 /sex="mat A"
 /dev_stage="Aggregation stage"

BASE COUNT 116 a 58 c 52 g 139 t

ORIGIN

Alignment Scores:
 Pred. No.: 363 Length: 365
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x B0326257 (1-365)

QY 123 GLYAspleuHISProThrHis 130
 |||||||
 Db 168 GGTGATTTCATCCACTACTCAT 191

RESULT 26
 A0191207 366 bp DNA linear GSS 01-NOV-1998
 LOCUS
 DEFINITION HS.2195_B2_H01_MR CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=2195 Col=2 Row=P, DNA sequence.

ACCESSION A0191207
 VERSION A0191207.1 GI:3583849
 KEYWORDS GSS.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 366)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Ketter,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Title Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

JOURNAL MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2195 row: P column: 2
 Class: BAC ends
 High quality sequence stop: 366.
 Location/Qualifiers
 1..366
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2195 Col=2 Row=P"
 /clone_1id="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
 E-Coli DH10B"

BASE COUNT 163 a 64 c 55 g 83 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 364 Length: 366
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 17 Gaps: 0

US-10-024-955-7 (1-213) x A0191207 (1-366)

QY 70 GLUleuAlaMetArgAsnIleGlu 77
 |||||||
 Db 12 GAACAGCAATGCGAACATGCA 35

RESULT 27
 A0323927 367 bp DNA linear GSS 04-MAY-1999
 LOCUS
 DEFINITION RCI11-105N7.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-105N7,
 DNA sequence.

ACCESSION A0323927
 VERSION A0323927.1 GI:4051199
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 367)
 Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,
 Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
 Use of human BAC End Sequences for Sequence-Ready Map Building
 Unpublished (1998)
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetlgr.org
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
 Research Genetics (inforesgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..367
 /organism="Homo sapiens"
 /db_xref="GDB:7540254"
 /db_xref="taxon:9606"
 /clone="RPCI-11-105N7"
 /clone_1id="RPCI-11"
 /sex="male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC library"

BASE COUNT 141 a 72 c 67 g 87 t
 ORIGIN

Alignment Scores:
 Pred. No.: 366 Length: 367
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 17 Gaps: 0

US-10-024-955-7 (1-213) x A0323927 (1-367)

QY 106 GLYValHISAspIleValSer 113
 |||||||
 Db 301 GGGGTTTCATGATGATGATGCA 278

RESULT 28
 A1065613 368 bp mRNA linear EST 24-JUL-1998
 LOCUS
 DEFINITION ag90e10.x1 maize inflorescence immature ear library Zea mays cDNA
 clone ag90e10 3', mRNA sequence.

ACCESSION A1065613
 VERSION A1065613.1 GI:3341020
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 368)
 Schutz,K., de la Bastide,M., Gnoj,L., Habermann,K., Huang,E.N.,
 Parnell,L.D., Dedhia,N., Martienssen,R. and McCombie,W.R.
 Expressed sequence tags from Z. mays
 Unpublished (1998)
 Contact: W. Richard McCombie
 Lila Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884

Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: ag90 row: e column: 10
Seq primer: M13 forward universal -21
High quality sequence stop: 368.
Location/Qualifiers

FEATURES

source

1.368
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ag90e10"
/clone_lib="maize inflorescence immature ear library"
/sex="female"
/tissue_type="immature ear"
/note="vector: pBLUESCRIPT SK+ (X52325); Site_1: XhoI;
Site_2: EcoRI; This library is described in Schmidt, Hake,
et al., (1993) Plant Cell 5:729-737. cDNAs are
directionally cloned into the XhoI and EcoRI sites; XhoI
is near the polyA tail. Most reads from this library are
3' in direction. Additional information on this library as
well as ftp access to all sequences can be found at
http://www.cshl.org/maizegenome"

BASE COUNT 135 a 89 c 64 g 80 t
ORIGIN

Alignment Scores:

Pred. No.: 367 Length: 368
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x A1065613 (1-368)

QY 86 ArgGlnGlyAspAlaAsnValLys 93

Db 316 AGACAAGGCGATGCAATGTCAG 339

RESULT 29

BJ359974

LOCUS BJ359974 378 bp mRNA linear EST 07-MAR-2002
DEFINITION Dictyostellium discoideum cDNA library, CF Dictyostellium

ACCESSION BJ359974

VERSION BJ359974.1 GI:19259569

KEYWORDS

EST.

SOURCE

Dictyostellium discoideum.

REFERENCE Dictyostellium discoideum.

AUTHORS

TITLE

JOURNAL

Unpublished (2002)

Contact: Tadasu Shin-I

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

FEATURES

source

1.378
/organism="Dictyostellium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="ddc3114"

/clone_lib="Dictyostellium discoideum cDNA library, CF"

/sex="mat A"

/dev_stage="Culmination stage"

BASE COUNT 118 a 62 c 61 g 136 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 379 Length: 378
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ359974 (1-378)

QY 123 GlyAspLeuHisProThrHis 130

Db 134 GGTGATTTCATCCACTACTCAT 157

RESULT 30

BJ326200

LOCUS BJ326200 394 bp mRNA linear EST 05-MAR-2002
DEFINITION Dictyostellium discoideum cDNA library, AF Dictyostellium

ACCESSION BJ326200

VERSION BJ326200.1 GI:19156330

KEYWORDS

EST.

Dictyostellium discoideum.

Dictyostellium discoideum.

Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.

1 (bases 1 to 394)

Full length cDNA of Dictyostellium discoideum at the aggregation

stage

Contact: Tadasu Shin-I

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

FEATURES

source

1.394
/organism="Dictyostellium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="ddai5013"

/clone_lib="Dictyostellium discoideum cDNA library, AF"

/sex="mat A"

/dev_stage="Aggregation stage"

BASE COUNT

119 a 66 c 62 g 147 t

ORIGIN

Alignment Scores:

Pred. No.: 400 Length: 394
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ326200 (1-394)

QY 123 GlyAspLeuHisProThrHis 130

Db 179 GGTGATTTCATCCACTACTCAT 202

RESULT 31

BB734318

LOCUS BB734318 400 bp mRNA linear EST 15-OCT-2001
DEFINITION BB734318 RIKEN full-length enriched, 6 days neonate spleen Mus

ACCESSION BB734318

VERSION BB734318.1 GI:16133468

KEYWORDS

EST.
house mouse.

SOURCE


```

ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imocani,K., Ishii
Y., Ito,M., Kawai,D., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
TITLE
JOURNAL
COMMENT      Unpublished (2001)
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagl,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Matsubara,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse
tissues.
FEATURES
SOURCE
Location/Qualifiers
1..400
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F420008H18"
/clone_lib="RIKEN full-length enriched, 6 days neonate
spleen"
/tissue_type="spleen"
/dev_stage="6 days neonate"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5 embryos
and E12.5 female mesonephros/gonad) and one newborn
ovary cDNA library. Average insert size 1.5 kb. All
source libraries are cloned unidirectionally with Oligo(dT
)-Not primers. References include: (1) Genome-wide
expression profiling of mid-gestation placenta and embryo
using a 15,000 mouse developmental cDNA microarray, 2000,
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
Large-scale cDNA analysis reveals phased gene expression
patterns during preimplantation mouse development, 2000,
Development, 127: 1737-1749; (3) Genome-wide mapping of
unselected transcripts from extraembryonic tissue of
7.5-day mouse embryos reveals enrichment in the t-complex
and under-representation on the x chromosome, 1998, Hum
Mol Genet 7: 1967-1978."

BASE COUNT      66 a      129 c      111 g      94 t
ORIGIN
Alignment Scores:
Pred. No.:      408      Length:      400
Score:          8.00      Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    3.76%      Indels:      0
DB:            13      Gaps:      0
US-10-024-955-7 (1-213) x BJ388026 (1-400)
OY      123 GYASPLEUHisProthrHis 130
Db      163 GGTGATTTCATCCACTACTCAT 186
RESULT 33
LOCUS      BJ329427
DEFINITION BJ329427 Dictyostellium discoideum cDNA library, AF Dictyostellium
ACCESSION BJ329427
VERSION    BJ329427.1 GI:19159557
KEYWORDS   EST.

BASE COUNT      400 bp      mRNA      linear      EST 08-MAR-2002
dictyostellium discoideum cDNA library, SF Dictyostellium
discoidium cDNA clone dds7b13 5', mRNA sequence.
ORIGIN
Alignment Scores:
Pred. No.:      408      Length:      400
Score:          8.00      Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    3.76%      Indels:      0
DB:            13      Gaps:      0
US-10-024-955-7 (1-213) x BJ388026 (1-400)
OY      8 AlAlAlAlAlAlAlAlAlAlAl 15
Db      253 GCTGCTTTCCTTTCCTTCCTCG 276
RESULT 32
LOCUS      BJ388026
DEFINITION BJ388026 Dictyostellium discoideum cDNA library, SF Dictyostellium
discoidium cDNA clone dds7b13 5', mRNA sequence.
ACCESSION BJ388026
VERSION    BJ388026.1 GI:19297410
KEYWORDS   EST.
SOURCE      Dictyostellium discoideum.
ORGANISM    Eukaryota; Eukaryota; Dictyostellida; Dictyostellium.
REFERENCE   1 (bases 1 to 400)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE       Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshtn@genes.nig.ac.jp.
FEATURES
SOURCE
Location/Qualifiers
1..400
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds7b13"
/clone_lib="Dictyostellium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="slug stage"
BASE COUNT      127 a      65 c      60 g      147 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      408      Length:      400
Score:          8.00      Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    3.76%      Indels:      0
DB:            13      Gaps:      0
US-10-024-955-7 (1-213) x BJ388026 (1-400)
OY      123 GYASPLEUHisProthrHis 130
Db      163 GGTGATTTCATCCACTACTCAT 186
RESULT 33
LOCUS      BJ329427
DEFINITION BJ329427 Dictyostellium discoideum cDNA library, AF Dictyostellium
ACCESSION BJ329427
VERSION    BJ329427.1 GI:19159557
KEYWORDS   EST.

```

SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
AUTHORS 1 (bases 1 to 403)
TITLE Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the aggregation stage

JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
location/Qualifiers

FEATURES
source 1..403
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="ddad24n19"
/clone_lib="Dictyostellium discoideum cDNA library, AP"
/sex="mat A"
/dev_stage="Aggregation stage"

BASE COUNT 127 a 65 c 60 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 412 Length: 403
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ329427 (1-403)

QY 123 GlyAspleuhisProthrhrHs 130
Db 162 GGTGATTTCATCCCTACTCAT 185

RESULT 34
BM531053/c 406 bp mRNA linear EST 19-FEB-2002
LOCUS fw92a07.y1 Gong zebrafish ovary Danio rerio cDNA clone 5617812 5'
DEFINITION similar to TR:09QZB9 09QZB9 DYNACTIN SUBUNIT P25. ; mRNA sequence.
ACCESSION BM531053
VERSION BM531053.1 GI:18739380
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 406)
Clart, M., Johnson, S.L., Lebrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Teising, B., Allen, J., Beck, Y., Person, B.,
Smaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Rittler, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
The library was constructed by Dr. Z. Gong. DNA sequencing by:
Washington University Genome Sequencing Center St. Louis. Please
contact Zhiyuan Gong for further information on this library
(National University of Singapore, Department of Biological

Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: T3 ET from Amersham
High quality sequence stop: 394.
location/Qualifiers

FEATURES
source 1..406
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="5617812"
/clone_lib="Gong zebrafish ovary"
/sex="female"
/dev_stage="4-5 month"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary (pooled); Vector: pBluescript SK-;
Site_1: XhoI; Site_2: EcoRI; Poly A+ RNA was isolated from
the ovaries of 2 female adult zebrafish (4-5 month old).
cDNAs were made using oligo-dT primers and inserted into
lambda Zap II vector (Stratagene) by Dr. Z. Gong. In vivo
mass-excision to pBluescript SK- following the Washington
University protocol
(http://genome.wustl.edu/est/lambda-protocol.shtml).
Please contact Zhiyuan Gong for further information on
this library (National University of Singapore,
Department of Biological Sciences, Lower Kent Ridge Road,
Singapore 119260)."

BASE COUNT 102 a 99 c 99 g 106 t
ORIGIN

Alignment Scores:
Pred. No.: 415 Length: 406
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BM531053 (1-406)

QY 186 LeuThrAlaIlePheGlnAspThr 193
Db 241 CTTACAGCAATCTTCAAGACACA 218

RESULT 35
BJ391088 413 bp mRNA linear EST 08-MAR-2002
LOCUS BJ391088 Dictyostellium discoideum cDNA library, SF
DEFINITION dictyostellium cDNA clone dds15d01 5', mRNA sequence.
ACCESSION BJ391088
VERSION BJ391088.1 GI:19302174
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
1 (bases 1 to 413)
Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the slug stage
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
location/Qualifiers

FEATURES
source 1..413
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="dds15d01"
/clone_lib="Dictyostellium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"

BASE COUNT 125 a 67 c 71 g 150 t

ORIGIN

Alignment Scores:

Pred. No.:	425	Length:	413
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.76%	Indels:	0
DB:	13	Gaps:	0

US-10-024-955-7 (1-213) x BJ391088 (1-413)

QY 123 G1yAspleuHisProThrThHis 130
|||||

Db 133 GGTGATTGCATCCACTACTCAT 156

RESULT 36
BI200525/c 415 bp mRNA linear EST 10-JUL-2001
LOCUS n3el0fs.r1 Fusarium sporotrichioides Trl 10 overexpressed cDNA
DEFINITION library Fusarium sporotrichioides cDNA clone n3el0fs 5', mRNA
sequence.

ACCESSION BI200525
VERSION BI200525.1 GI:14666497
KEYWORDS EST.
SOURCE Fusarium sporotrichioides.
ORGANISM Fusarium sporotrichioides.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE 1 (bases 1 to 415)
AUTHORS Ren,Q., Tag,A., Peplov,A., Lai,H., Kupfer,C., Peterson,A., Beremand
M. and Roe,B.
TITLE Analysis of a Fusarium sporotrichioides EST database
JOURNAL Unpublished (2001)
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Contact Dr. Marian Beremand regarding clone availability. Included
is the best homolog from a blastx search of Genbank nr 04-09-01
[Homosapiens
Seq primer: T3
High quality sequence stop: 234.
Location/Qualifiers
1. .415
/organism="Fusarium sporotrichioides"
/strain="Trl 10"
/db_xref="taxon:5514"
/clone="n3el0fs"
/clone_lib="Fusarium sporotrichioides Trl 10 overexpressed
cDNA library"
/note="Vector: pBluescript SK-. Site.1: EcoRI; Site.2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT 92 a 157 c 93 g 73 t
ORIGIN

Alignment Scores:

Pred. No.:	427	Length:	415
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.76%	Indels:	0
DB:	13	Gaps:	0

US-10-024-955-7 (1-213) x BI200525 (1-415)

QY 140 ValAlaLeuSerLeuGluIleSer 147
|||||

Db 294 GTGCACCTCTCGTCGAGATCAGT 271

RESULT 37
BI418268 415 bp mRNA linear EST 15-AUG-2001
LOCUS lJNEST46h1r Lotus japonicus node library 5 and 7 week-old Lotus
DEFINITION lJNEST46h1r Lotus japonicus node library 5 and 7 week-old Lotus
japonicus cDNA 5', mRNA sequence.
ACCESSION BI418268
VERSION BI418268.1 GI:15189291
KEYWORDS EST.
SOURCE Lotus japonicus.
ORGANISM Lotus japonicus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.

REFERENCE 1 (bases 1 to 415)
AUTHORS Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
TITLE Lotus japonicus root node ESTs: tools for functional genomics
JOURNAL Unpublished (2000)
COMMENT Contact: Udvardi MK
Molecular Plant Nutrition
Max Planck Institute of Molecular Plant Physiology
Am Muehlenberg 1, 14476 Golm, Germany
Fax: 49 331 567 8250
Email: udvardi@mpi-np-golm.mpg.de
Seq primer: T7
High quality sequence stop: 415.
Location/Qualifiers
1. .415
/organism="Lotus japonicus"
/cultivar="Gifu (B-129)"
/db_xref="taxon:34305"
/clone_lib="Lotus japonicus node library 5 and 7
week-old"
/dev_stage="5 and 7 week-old plants"
/note="Organ: Node; Vector: pSPORT1; Site.1: SalI;
Site.2: NotI; The library was prepared using mRNA
extracted from nodules of 5 and 7 week-old Lotus plants.
Nodules were induced by, and contained Mesorhizobium
strain R7A."

BASE COUNT 141 a 79 c 47 g 148 t
ORIGIN

Alignment Scores:

Pred. No.:	427	Length:	415
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.76%	Indels:	0
DB:	13	Gaps:	0

US-10-024-955-7 (1-213) x BI418268 (1-415)

QY 172 leuSerIleLeuAspProIlePhe 179
|||||

Db 144 CTCTCAATCTGATCCTATCTTT 167

RESULT 38
BJ391085 416 bp mRNA linear EST 08-MAR-2002
LOCUS BJ391085
DEFINITION BJ391085 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds15c01 5', mRNA sequence.
ACCESSION BJ391085
VERSION BJ391085.1 GI:19302171
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
REFERENCE 1 (bases 1 to 416)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)

/ tad_mosc- dn10b (ampicillin resistant)

Pred. No.:	440	422
Score:	8.00	8
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	3.76%	Indels: 0
BB:	13	Gaps: 0

US-10-024-955-7 (1-213) x BJ388944 (1-425)

OY 123 GLYASPLEUHISPROTHrThHis 130
|||||

DB 140 GGTGATTTCATCCCACTACTCAT 163

RESULT 41

LOCUS

BJ386722 438 bp mRNA linear EST 08-MAR-2002
Dictyostelium discoidium cDNA library, SF Dictyostelium

ACCESSION BJ386722
dictyostelium cDNA clone dds10f11 5', mRNA sequence.

VERSION BJ386722.1 GI:19296106

KEYWORDS EST.

SOURCE

Dictyostelium discoidium.

ORGANISM Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE 1 (bases 1 to 438)

AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.

TITLE Full length cDNA of Dictyostelium discoidium at the slug stage

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

Location/Qualifiers

FEATURES

source

1..438

/organism="Dictyostelium discoidium"

/strain="AX4"

/db_xref="taxon:4689"

/clone="dds10f11"

/clone_lib="Dictyostelium discoidium cDNA library, SF"

/sex="mat A"

/dev_stage="slug stage"

BASE COUNT 132 a 78 c 81 g 147 t

ORIGIN

Alignment Scores:

Pred. No.: 457 Length: 438

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.76% Indels: 0

DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ386722 (1-438)

OY 123 GLYASPLEUHISPROTHrThHis 130
|||||

DB 126 GGTGATTTCATCCCACTACTCAT 149

RESULT 42

LOCUS

BJ388417 441 bp mRNA linear EST 08-MAR-2002
Dictyostelium discoidium cDNA library, SF Dictyostelium

ACCESSION BJ388417
dictyostelium cDNA clone dds9p09 5', mRNA sequence.

VERSION BJ388417.1 GI:19299503

KEYWORDS EST.

SOURCE Dictyostelium discoidium.

ORGANISM

Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE 1 (bases 1 to 441)

AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.

TITLE Full length cDNA of Dictyostelium discoidium at the slug stage

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES

source

1..441

/organism="Dictyostelium discoidium"

/strain="AX4"

/db_xref="taxon:4689"

/clone="dds9p09"

/clone_lib="Dictyostelium discoidium cDNA library, SF"

/sex="mat A"

/dev_stage="slug stage"

BASE COUNT 134 a 70 c 70 g 167 t

ORIGIN

Alignment Scores:

Pred. No.: 461 Length: 441

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.76% Indels: 0

DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ388417 (1-441)

OY 123 GLYASPLEUHISPROTHrThHis 130
|||||

DB 162 GGTGATTTCATCCCACTACTCAT 185

RESULT 43

LOCUS

BJ358769 442 bp mRNA linear EST 07-MAR-2002
Dictyostelium discoidium cDNA library, CF Dictyostelium

ACCESSION BJ358769
dictyostelium cDNA clone ddc1le07 5', mRNA sequence.

VERSION BJ358769

KEYWORDS EST.

SOURCE Dictyostelium discoidium.

ORGANISM

Dictyostelium discoidium.

Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE 1 (bases 1 to 442)

AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.

TITLE Full length cDNA of Dictyostelium discoidium at the culmination

JOURNAL stage Unpublished (2002)

COMMENT Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

Location/Qualifiers

FEATURES

source

1..442

/organism="Dictyostelium discoidium"

/strain="AX4"

/db_xref="taxon:4689"

/clone="ddc1le07"

/clone_lib="Dictyostelium discoidium cDNA library, CF"

/sex="mat A"

/dev_stage="culmination stage"

BASE COUNT 134 a 70 c 71 g 167 t

ORIGIN

Alignment Scores:

Pred. No.: 463 Length: 442

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.76% Indels: 0

DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ358769 (1-442)

OY 123 GtAspLeuHisProthrThHis 130
 DB 162 GGTGATTTCATCCACTACTCAT 185

RESULT 44
 BQ146887/c 443 bp mRNA linear EST 24-APR-2002

LOCUS BQ146887
 DEFINITION NF029D03FL1029 Developing flower Medicago truncatula cDNA clone
 NF029D03FL 5', mRNA sequence.

ACCESSION BQ146887
 VERSION BQ146887.1 GI:20283946

KEYWORDS EST.

SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 1 (bases 1 to 443)
 Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
 Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
 Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula flower library
 Unpublished (2001)
 JOURNAL Contact: May GD
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7391
 Fax: 580 221 7380
 Email: gdmay@noble.org
 Insert Length: 443 Std Error: 0.00
 Plate: 029 Row: D Column: 03
 Seq primer: TCACACAGCAACAGCTATGAC.

FEATURES
 Source Location/Qualifiers
 1..443
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF029D03FL"
 /clone_lib="Developing flower"
 /tissue_type="Developing flowers"
 /dev_stage="Developmentally pooled. Contains a mixture of
 very young, developing, fully-opened flowers and flowers
 in early transition into pods."
 /note="Vector: Lambda Zap; cDNA was prepared from polyA+
 enriched, pooled samples of equivalent amounts of total
 RNA from very young, developing, fully-opened flowers and
 flowers transitioning into pods. The cDNA was
 directionally ligated into the Uni-Zap XR vector
 (Stratagene) and packaged using the Gigapack III Gold
 packaging extracts. Phagemids containing cDNA inserts were
 in vivo excised from the recombinant Uni-Zap XR vector
 using ExAssist helper phage and the E. coli strain
 XL1-Blue MRF' (Stratagene). Excised plasmids were plated
 using SOLR cells."

BASE COUNT 162 a 76 c 100 g 105 t
 ORIGIN

Alignment Scores: 464 Length: 443
 Pred. No.: 8.00 Matches: 8
 Score: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 14 Gaps: 0

US-10-024-955-7 (1-213) x BQ146887 (1-443)

OY 178 lIepheGLyValIeuSerAspVal 185
 DB 432 ATCTTCGGTGTCTTATCAAGATGTT 409

RESULT 45
 BQ358897 445 bp mRNA linear EST 07-MAR-2002

LOCUS BQ358897
 DEFINITION BQ358897 Dictyostellium discoideum cDNA library. Cf Dictyostellium
 discoideum cDNA clone ddc1119 5', mRNA sequence.

ACCESSION BQ358897
 VERSION BQ358897.1 GI:19258492

KEYWORDS EST.

SOURCE Dictyostellium discoideum.
 ORGANISM Dictyostellium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 1 (bases 1 to 445)
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 Full length cDNA of Dictyostellium discoideum at the culmination
 stage
 Unpublished (2002)
 JOURNAL Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 Source Location/Qualifiers
 1..445
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="ddc1119"
 /clone_lib="Dictyostellium discoideum cDNA library, Cf"
 /sex="mat A"
 /dev_stage="Culmination stage"

BASE COUNT 135 a 70 c 71 g 169 t
 ORIGIN

Alignment Scores: 467 Length: 445
 Pred. No.: 8.00 Matches: 8
 Score: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.76% Indels: 0
 DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BQ358897 (1-445)

OY 123 GtAspLeuHisProthrThHis 130
 DB 165 GGTGATTTCATCCACTACTCAT 188

RESULT 46
 BQ362311 446 bp mRNA linear EST 07-MAR-2002

LOCUS BQ362311
 DEFINITION BQ362311 Dictyostellium discoideum cDNA library. Cf Dictyostellium
 discoideum cDNA clone ddc21c10 5', mRNA sequence.

ACCESSION BQ362311
 VERSION BQ362311.1 GI:19261906

KEYWORDS EST.

SOURCE Dictyostellium discoideum.
 ORGANISM Dictyostellium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 1 (bases 1 to 446)
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 Full length cDNA of Dictyostellium discoideum at the culmination
 stage
 Unpublished (2002)
 JOURNAL Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 Location/Qualifiers

```

source
1. .446
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44699"
/clone_id="ddc2ic10"
/dev_stage="Culmination stage"
BASE COUNT 135 a 72 c 71 g 168 t
ORIGIN

Alignment Scores:
Pred. No.: 468 Length: 446
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 13 Gaps: 0

US-10-024-955-7 (1-213) x BJ362311 (1-446)

Oy 123 GYASPLEUHSPTHTHRHS 130
Db 162 GGTGATTTCATCCACTACTCAT 185

RESULT 47
AI066202 450 bp mRNA linear EST 21-SEP-2000
LOCUS TEN02317 T. cruzi epimastigote normalized cDNA library Trypanosoma
ACCESSION AI066202
VERSION AI066202.1 GI:3366367
KEYWORDS EST.
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE Porcel,B.M., Tran,A.-N., Tammi,M., Nyarady,Z., Rydaker,M., Urmenyi
AUTHORS T.P., Roodinelli,E., Petersson,U., Andersson,B. and Aslund,L.
TITLE Gene survey of the pathogenic protozoan Trypanosoma cruzi
JOURNALS Genome Res. 10 (8), 1103-1107 (2000)
MEDLINE 20414748
COMMENT Contact: Aslund L
Department of Medical Genetics
Uppsala University
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden
Tel: 46 18 471 45 85
Fax: 46 18 52 68 49
Email: lena.aslund@medgen.uu.se
Seq primer: T7
High quality sequence stop: 450.
Location/Qualifiers
FEATURES
Source
1. .450
/organism="Trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone_id="25e12"
/clone_lib="T. cruzi epimastigote normalized cDNA library"
/cell_type="epimastigote"
/note="cDNA library constructed with oligo dT primed
epimastigote mRNA and cloned in pETc18D phagemid with
modified polylinker (PHARMACIA)"
BASE COUNT 69 a 89 c 110 g 182 t
ORIGIN

Alignment Scores:
Pred. No.: 473 Length: 450
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 9 Gaps: 0

```

```

US-10-024-955-7 (1-213) x AI066202 (1-450)
Oy 180 GYVALLUHSRASPALLEUHR 187
Db 20 GGAGTCTTCTGACCTACTACT 43

RESULT 48
AO353113 450 bp DNA linear GSS 24-JAN-1999
LOCUS CITBI-EI-2540H3.TF CITBI-EI Homo sapiens genomic clone 2540H3, DNA
DEFINITION sequence.
ACCESSION AO353113
VERSION AO353113.1 GI:4180448
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
AUTHORS Venter,J.C.
TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
JOURNAL Map Building
COMMENT Unpublished (1997)
Other GSSs: CITBI-EI-2540H3.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (Info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
Source
1. .450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="2540H3"
/clone_lib="CITBI-EI"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 138 a 69 c 102 g 141 t
ORIGIN

Alignment Scores:
Pred. No.: 473 Length: 450
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 17 Gaps: 0

US-10-024-955-7 (1-213) x AO353113 (1-450)
Oy 79 ARGGLYLEULSGLMETYSARG 86
Db 393 AGAGGGCTCAGCAGTCAAGAGG 416

RESULT 49
BG660965 451 bp mRNA linear EST 11-MAY-2001
LOCUS TGESTzya12b10.y1 TgyEG118 Tachyzoite cDNA library Toxoplasma gondii
DEFINITION cdna clone TGESTzya12b10.y1 5', mRNA sequence.
ACCESSION BG660965
VERSION BG660965.1 GI:13805056
KEYWORDS EST.

```

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Toxoplasma gondii.
Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 451)
Tang, K., Cole, R., Fogarty, S., Sibley, L.D., Ajlaka, J.A., White, M.,
Clifton, S., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M.,
Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter,
E., Bennett, J., Franklin, C., Tsagaris, J., Ronko, I., Kennedy,
S., Maguire, L., Waterson, R. and Wilson, R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.
Toxoplasma EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxowatson.wustl.edu
Contact David Sibley (toxowest@orcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 415.

FEATURES
source
Location/Qualifiers
1..451
/organism="Toxoplasma gondii"
/strain="VEG"
/db_xref="taxon:5811"
/clone="cgscstzyl2p10.y1"
/clone_lib="TYVEG18 Tachyzoite cDNA library"
/dev_stage="Tachyzoite"
/lab_host="DH10B"
/note="Vector: pBluescript SK; Site:1: EcoRI; Site:2: XhoI
/note="This library was constructed by Keliang Tang, Robert
Cole, and L. David Sibley at Washington University. CDNA
were synthesized from poly(A)+ RNA by oligod(T) priming,
size-selected and directionally cloned into the Uni-ZAP XR
lambda vector (Stratagene). The primary library was mass
excised as phagemids and rescued in SOLR cells. The
plasmid library was recovered from the SOLR cells and
transformed in mass into DH10B (GeneHog, Research Genetics
, Inc.) for sequencing. WARNING: This library may contain
a small percentage contaminants from human fibroblast
cells."

BASE COUNT
ORIGIN
102 a 127 c 118 g 104 t

Alignment Scores:
Pred. No.: 475 Length: 451
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 12 Gaps: 0

US-10-024-955-7 (1-213) x BG660965 (1-451)

QY 180 GYValleuseraspvalleuthr 187
Db 23 GGAGTGTGAGCATGTTTGACC 46

RESULT 50
AI946856/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

455 bp mRNA linear EST 08-JAN-2001
bs31h08.y1 Drosophila melanogaster adult testis library Drosophila
melanogaster cDNA clone bs31h08 5', mRNA sequence.
AI946856
AI946856.2 GI:9992170
EST.
fruit fly,
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 455)
Andrews, J., Bouffard, G.G., Cheadle, C., Lu, J., Becker, K.G. and
Oliver, B.
Gene discovery using computational and microarray analysis of
transcription in the drosophila melanogaster testis
Genome Res. 10 (12), 2030-2043 (2000)
20568492
On Aug 17, 1999 this sequence version replaced gi:5737282.
Contact: Brian Oliver
Laboratory of Cellular and Developmental Biology
NIDDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239
Email: Oliver@helix.nih.gov,
http://www.nidk.nih.gov/intram/people/boliver.htm
Tissue isolation and library construction performed at the National
Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
http://www.nidk.nih.gov/intram/people/boliver.htm). DNA sequencing
and analyses performed by National Institutes of Health Intramural
Sequencing Center (NISC; see http://www.nisc.nih.gov).
Plate: 31 row: h column: 08
Seq primer: M13RP1 reverse primer (AB1).

FEATURES
source
Location/Qualifiers
1..455
/organism="Drosophila melanogaster"
/strain="y[*] w[67c1]/y"
/db_xref="taxon:7227"
/clone="bs31h08"
/clone_lib="Drosophila melanogaster adult testis library"
/sex="male"
/dev_stage="1-5 day adult"
/lab_host="SOLR (Stratagene)"
/note="Organ: testis; Vector: pBluescript SK (Stratagene);
Site:1: EcoR I; Site:2: Xho I; Testes dissected from 1-5
day adult y[*] w[67c1]/y males raised at 25°C. RNA
isolated using Trizol (Life Technologies) and a single
round of Poly(A)+ selection using Oligotex (Qiagen). cDNA
library constructed using Stratagene ZAP-cDNA synthesis
kit. Oligo dT-primed, size fractionated ~1-6 kb, and
directionally cloned at EcoR and XhoI in Uni-ZAP XR.
Following a single round of amplification pBluescript SK
phagemids were mass excised. A distribution channel for
clones is being sought, but not currently available.
Requests for clones cannot be honored."

BASE COUNT
ORIGIN
149 a 108 c 117 g 81 t

Alignment Scores:
Pred. No.: 480 Length: 455
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.76% Indels: 0
DB: 9 Gaps: 0

US-10-024-955-7 (1-213) x AI946856 (1-455)

QY 8 AAlaValaAlaPheValaAlaVal 15
Db 354 GCTGCTTGCTTCCTTGTGCTGTT 331

Search completed: February 20, 2003, 19:25:10
Job time: 1482 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 14:34:26 ; Search time 2012 seconds

(without alignments)
3080.963 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 1068

Sequence: 1 MMKFLIAAFAVAVSADPI.....VRKEMTKVLAFAFKRELEKN 213

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_{p2n}.model -DEV=xlp
-O=/cgn2.1/USPFO.spool/US10024955/runat_14022003_130658_23705/app-query.fasta.1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptio -NORMEXT=HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10024955 -EGCN_1.1_3745 -runat_14022003_130658_23705 -MCPU=6 -ICPU=3
-NO_XLPTX -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mus:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1068	100.0	761	3	S80655
2	1068	100.0	761	6	AX028702
3	962	90.1	812	3	DP037044
4	959	89.8	812	6	AR098569
5	959	89.8	812	6	AX028697
6	251.5	23.5	767	3	LDE271058
7	104.5	9.8	103996	8	ATF8M21
8	103	9.6	3737	1	AB016077
9	96.5	9.0	11862	1	AE005025
10	96	9.0	892	6	AX350829
11	96	9.0	892	6	AX350831
12	96	9.0	893	6	AX350827
13	96	9.0	1198	6	BC034696
14	96	9.0	1233	6	AX430414
15	96	9.0	23266	2	AC026300
16	96	9.0	178012	2	AC131214
17	96	9.0	181780	2	AC117474
18	96	9.0	207698	2	AC112171
19	95.5	8.9	558	1	AF437087
20	95.5	8.9	315079	1	MPULM03
21	95	8.9	158714	8	AC082644
22	93.5	8.8	1266	8	ASNGALIA
23	93.5	8.8	293250	1	CNSPAX02
24	93.5	8.8	349980	6	AX041918
25	92.5	8.7	10558	1	AE010403
26	92.5	8.7	40909	1	SCI35
27	92.5	8.7	89934	8	AC007067
28	92	8.6	558	1	AF436928
29	92	8.6	558	1	AF436929
30	92	8.6	558	1	AF436936
31	92	8.6	558	1	AF436950
32	92	8.6	558	1	AF436963
33	92	8.6	558	1	AF436980
34	92	8.6	558	1	AF437078
35	92	8.6	558	1	AF437182
36	92	8.6	558	1	AF437184
37	92	8.6	558	1	AF437251
38	92	8.6	558	1	AF437285
39	91.5	8.6	10142	1	TARPOG
40	91.5	8.6	10376	1	AE002228
41	91.5	8.6	10378	1	AE001593
42	91.5	8.6	299650	1	AP002545
43	91.5	8.6	338100	1	TACID2
44	91	8.5	558	1	AF436987
45	91	8.5	558	1	AF437185

RESULT 1

ALIGNMENTS

S80655
 LOCUS 580655 761 bp mRNA linear INV 10-MAY-1996
 DEFINITION allergen Der f 7 (Dermatophagoides pteronyssinus-house dust mites, mRNA, 761 nt).
 ACCESSION S80655
 VERSION S80655.1 GI:1311688
 KEYWORDS
 SOURCE Dermatophagoides pteronyssinus.
 ORGANISM Dermatophagoides pteronyssinus
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae; Dermatophagoides.
 REFERENCE 1 (bases 1 to 761)
 AUTHORS Shen,H.D., Chua,K.Y., Lin,W.L., Hsieh,K.H. and Thomas,W.R.
 TITLE Molecular cloning and immunological characterization of the house dust mite allergen Der f 7
 JOURNAL Clin. Exp. Allergy 25 (10), 1000-1006 (1995)
 MEDLINE 96120794
 PUBMED 8556554
 REMARK Genbank staff at the National Library of Medicine created this entry [NCBI gtblseq 1/4/94] from the original journal article. This sequence comes from Fig. 1.
 FEATURES
 source 1..761
 /organism="Dermatophagoides pteronyssinus"
 /db_xref="taxon:6956"
 gene 1..761
 /gene="allergen Der f 7"
 CDS 43..684
 /gene="allergen Der f 7"
 /note="This sequence comes from Fig. 1"
 /codon_start=1
 /product="allergen Der f 7"
 /protein_id="AAB35977.1"
 /db_xref="GI:1311689"
 /translation="MKKFLIAVAFAVAVSADPIHYDKITEERINKAIDDAIAIEOSE
 TIDPKVDPHADKFERHVGIDFKGLARNLKARGLKQKMGDANVGEIGYKAH
 LIGVHDDIVSMEDLAYKLGIDHPHTTHVISTIDPVALSLEISDEGNTMTSEVR
 OFANVNNHGGISILDPIFGVSDVLTALFQDTVRKMKVLAFAKRELEKN"
 BASE COUNT 245 a 134 c 137 g 245 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.01e-83 Length: 761
 Score: 1068.00 Matches: 213
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0
 US-10-024-955-7 (1-213) x S80655 (1-761)
 QY 1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
 DB 43 ATGATGAATTTTGTGATGCTGCGCGTGCATTTTGCCTTGGCTGATCCAAAT 102
 QY 21 HisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAlaIle 40
 DB 103 CACTATGATTAATTCACCGAAGAAATCAACAAGCTATTGATGATGCTGCTATT 162
 QY 41 GluGlnSerGluThrIleAspPrometLysValProAspHisAlaAspLysPheGluArg 60
 DB 163 GAACAATCCGAACAATAGTCCAAATGAAGTACTGATATCCCGCATTAATTCGAACGT 222
 QY 61 HisValGlyIleValAspPheLysGlyLeuAlaMetArgAsnIleGluAlaArgGly 80
 DB 223 CATGTTGGATTGTGATTTCAAAGGTGAATTAGCCATGCGAAACATTGAGGCTCGAGGA 282
 QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys 100
 DB 283 TTGAACAATTAAGACGTCAAGGTGATGCTAAATGTCAAAGGTGAAGGGGTATTGTTAAA 342
 QY 101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120

DB 343 GCTCATTTGTTGATCGGTGTTCCAGATGATATGCTCCATGGAATATCATTTAGCATAC 402
 QY 121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValVal 140
 DB 403 AAATGGGTGATCTTCATCCAACTCACTCACTGATTCGATATTCGAATTTGTGTT 462
 QY 141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArg 160
 DB 463 GCCCTTCCCTTGAATTTCTGATGAAGTAACATGATGATCACTTTTGAAGTAGCA 522
 QY 161 GlnPheAlaAsnValValAsnHisIleGlyLeuSerIleLeuAspProIlePheGly 180
 DB 523 CAATTCGCTAATTTGTCAACCATATTTGGTGGCTTTCAATCTTGATCCATTTTGGC 582
 QY 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys 200
 DB 583 GTTTATTCGATGATATGACCGCTATTTTCCAGACACCGTAGCTAAGAAATGACCAAA 642
 QY 201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
 DB 643 GTATTGGACACCATTTAAACGTGAATTGGAAAAAAT 681
 RESULT 2
 AX028702 761 bp DNA linear PAT 16-SEP-2000
 LOCUS
 DEFINITION Sequence 6 from Patent EP1018550.
 ACCESSION AX028702
 VERSION AX028702.1 GI:10189815
 KEYWORDS
 SOURCE European house dust mite.
 ORGANISM Dermatophagoides pteronyssinus
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae; Dermatophagoides.
 REFERENCE 1 (bases 1 to 761)
 AUTHORS Thomas,W.R. and Chua,K.Y.
 TITLE Allergenic protein and peptides from house dust mite and uses thereof
 JOURNAL Patent: EP 1018550-A 6 12-JUL-2000;
 INST CHILD HEALTH RESEARCH (AU)
 FEATURES
 source 1..761
 /organism="Dermatophagoides pteronyssinus"
 /db_xref="taxon:6956"
 CDS 43..684
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC09235.1"
 /db_xref="GI:10189816"
 /translation="MKKFLIAVAFAVAVSADPIHYDKITEERINKAIDDAIAIEOSE
 TIDPKVDPHADKFERHVGIDFKGLARNLKARGLKQKMGDANVGEIGYKAH
 LIGVHDDIVSMEDLAYKLGIDHPHTTHVISTIDPVALSLEISDEGNTMTSEVR
 OFANVNNHGGISILDPIFGVSDVLTALFQDTVRKMKVLAFAKRELEKN"
 BASE COUNT 245 a 134 c 137 g 245 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.01e-83 Length: 761
 Score: 1068.00 Matches: 213
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0
 US-10-024-955-7 (1-213) x AX028702 (1-761)
 QY 1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
 DB 43 ATGATGAATTTTGTGATGCTGCGCGTGCATTTTGCCTTGGCTGATCCAAAT 102
 QY 21 HisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAlaIle 40

Dd	103	CACCTATGATAAAANTCACCAGAAATCAACAAGAAGTATTGTATGATCCATTGGCTGTAATT	162
Oy	41	GluGlnSerGluThrIleAspPromeLeuValProASPHisAlaAspLysPheGluArg	60
Dd	163	GAAACAATCGAAGAACATATGATTCATTCGAAGACTGATCATGCCGATTAATTCGAACGT	222
Oy	61	HISVALIGLYILEVALASPPHELYSGLYGLEUALAMETARGASNILEGLNALAARGGLY	80
Dd	223	CATGTTGGTATTTGTGCATTTTCCAAAGGTAATTAGCCATGCGAACAATTGAGGCTCGAGGA	282
Oy	81	LeuLysGlnMetLysArgGlnGlyAspRlaAsnValLysGlyGlnGlnGlyIleValLys	100
Dd	283	TTGAAACAATGTAACACGTCAAGCTGATCTAATGTCCAAGGGAAGAGGATATGGTTAAA	342
Oy	101	ALAHLISLEULEULIEGLIVALHISASPSPILEVALSERMETGLUTYRASPLEUALTYR	120
Dd	343	GCTCATTTTGTTGATGCGGTTCACGATGATATACGCTCGATGAGATATGATTAGCATAC	402
Oy	121	LysLeuGlyAspLeuHisProthrThrlsVALIIIESERAPILEGINASPPHEVALVAL	140
Dd	403	AAATGGGTGATCTTCATCAACCAACCTCATGTCATTTGGATATTCAGATTTTGTGTT	462
Oy	141	AlaLeuSerLeuGluLisSerAspGlnGlyAsnIleThrMetThrSerPheGluValArg	160
Dd	463	GCTTGTCCTCGTAAGAAATTTGTATGATCAAGTAACATTAACATGACATCTTTGCAACTACGA	522
Oy	161	GlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly	180
Dd	523	CAATTCCGTAAATGTTGTCAACCATATTTGGTGTCTTTCAATCTTGGAATCCAAATTTTTGGC	582
Oy	181	ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys	200
Dd	583	GTATTATCTGATGTGTATGACCGCTATTTTCCAAAGCACCGTCGTAAGAAATGACCAA	642
Oy	201	ValLeuAlaProAlaPheLysArgLuuLeuGluLysAsn	213
Dd	643	GTATTGGCACCGACATTTAAACGTGAATTTGAAAAAAAT	681
RESULT 3			
LOCUS	DPU37044	812 bp	mRNA linear INV 31-OCT-1995
DEFINITION	Dermatophagoides pteronyssinus Der p 7 allergen polypeptide mRNA,		
ACCESSION	U37044		complete cds.
VERSION	U37044.1		GI:1045601
KEYWORDS			.
SOURCE	ORGANISM		Dermatophagoides pteronyssinus.
			Dermatophagoides pteronyssinus
			Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
			Acariformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;
			Dermatophagoides.
REFERENCE	AUTHORS		1 (bases 1 to 812)
	TITLE		Shen,H.D., Chu,K.Y., Lin,K.L., Hsieh,K.H. and Thomas,W.R.
			Molecular cloning of a house dust mite allergen with common
			antibody binding specificities with multiple components in mite
			extracts
JOURNAL	MEDLINE		Clin. Exp. Allergy 23 (11), 934-940 (1993)
PUBMED	20239583		
REFERENCE	10779281		
AUTHORS	2 (bases 1 to 812)		
TITLE	Thomas,W.R.		
JOURNAL	Direct Submission		
	Submitted (26-SEP-1995) Wayne R. Thomas, Molecular Biology		
	Division, Institute for Child Health Research, Princess Margaret		
	Hospital, Roberts Road, Subiaco, Western Australia, 6008, Australia		
FEATURES	Location/Qualifiers		
source	1..812		/organism="Dermatophagoides pteronyssinus"
			/db_xref="taxon:6956"
			/clone="HD6"
CDS	68..715		/codon_start=1
			/product="Der p 7 allergen polypeptide"

[illegible]

DEFINITION	sequence 1 from patent US 6077517.
ACCESSION	AR098569
VERSION	AR098569.1 GI:12808335
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 812)
AUTHORS	Thomas,W.Robert., and Chua,K.-Y.
TITLE	House dust mite allergen, Der p VII, and uses thereof
JOURNAL	Patent: US 6077517-A 1 20-JUN-2000;
FEATURES	Location/Qualifiers
Source	1..812
BASE COUNT	270 a /organism="unknown"
ORIGIN	133 c 142 g 267 t
Alignment Scores:	
Pred. No.:	2.78e-74 Length: 812
Score:	959.00 Matches: 182
Percent Similarity:	94.37% Conservative: 19
Best Local Similarity:	85.45% Mismatches: 12
Query Match:	89.79% Indels: 0
DB:	Gaps: 0
US-10-024-955-7 (1-213) x AR098569 (1-812)	
Oy	1 MetMetlyPheLeuLeuIAlAlaValAlaPheValAlaIleSerAlaAspProIle 20
Dd	68 ATGATGAATATTATTTATTTATTTATTCCTGCCAGACTTTGTGGCCGTTTGCGTGCATCAATT 127
Oy	21 HisTryAspLysIleThrGlulIleasnLysAlaIleAspAlaIleAlaIle 40
Dd	128 CACTATGATTAATAATCACCAGAATAATTACAAGCTGTGATGAAGCCGTCGTCAAATT 187
Oy	41 GluInSerGIuThrIleAspPrometLysValTroAspHisAlaAspLysPheGluArg 60
Dd	188 GAAAAATCCGAACCTTGATGCATGAAAGGTACCAGCATCTTGTGATTAATTGAAACGA 247
Oy	61 HisValGIyleValAspPheLysGIuIleuAlaMetArgAsnIleGluAlaArgLy 80
Dd	248 CATATTGGTATCATCGATTTTAAAGGTCAATTAAACATGCCAACAATTCAAGTTGAGGA 307
Oy	81 LeuLysGIuMetLysArgGlnGIuLysAspAlaAsnValLysGIuGIuGlnGIuIleValLys 100
Dd	308 TTAATAACAATGAACACGTGATGGATGCTTAATGTGAAGAAGTAAGATGGTGTGCAA 367
Oy	101 AlaHisLeuLeuIleGlyValHisAspSpIleValSerMetGIuTryAspLeuAlaTyr 120
Dd	368 GCCTATTGTGTGGTCGGTTCATGATGACCTGTTCCTTCACAAAGAAATATGATTTAGCATAC 427
Oy	121 LysLeuGIuLysPheHisProThrThriHisValIleSerAspIleGlnAspPheVal 140
Dd	428 AAATTGGGTGATCTTCATCCAAACACTCATATTCATTTGGGATATTCAGGATTTGTGTGC 487
Oy	141 AlaLeuSerLeuGluIleSerAspGIuLysAsnIleThrMetTrsPheGluValArg 160
Dd	488 GAATTATCGCTCGAAGTTAGCGAAGAAGGTAAATATGACATTCAGACATCGTTGGAAGTAGCT 547
Oy	161 GlnPheAlaAsnValValAsnHisIleGlyLysLeuSerIleLeuAspProIlePheGly 180
Dd	548 CAATTGGCAATGTTGTCAATCAATATGTGTTCCTTCATATTTGGATTCAAATTTGTGCT 607
Oy	181 ValLeuSerAspValLeuThrAlaIlePheGlnAspPhrValArgLysGIuMetThrLys 200
Dd	608 GTCTTATTCGATGTTTATCACCGCTATATTTTCCAGGATACCGTACGTGAGAAATGACCAA 667
Oy	201 ValLeuAlaProAlaPheLysArgGIuLeuGIuLysAsn 213
Dd	668 GTATTGGCACACGATTCAAAANAAGAAATTTGGAACGAAAC 706
RESULT 5	
AA028697	

LOCUS	AX028697	812 bp	DNA	linear	PAT 16-SEP-2000
DEFINITION	Sequence 1 from Patent EP1018550.				
ACCESSION	AX028697				
VERSION	AX028697.1	GI:10189810			
KEYWORDS					
SOURCE					
ORGANISM	European house dust mite.				
	Dermatophagoides pteronyssinus				
	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;				
	Acariiformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;				
	Dermatophagoides.				
	1 (bases 1 to 812)				
REFERENCE	Thomas, W.R. and Chua, K.Y.				
AUTHORS	Allergenic protein and peptides from house dust mite and uses				
TITLE	thereof				
JOURNAL	Patent: EP 1018550-A 1 12-JUL-2000;				
	INST CHILD HEALTH RESEARCH (AU)				
FEATURES	Location/Qualifiers				
source	1..812				
	/organism="Dermatophagoides pteronyssinus"				
	/db_xref="taxon:6956"				
CDS	68..715				
	/note="unamed protein product"				
	/codon_start=1				
	/protein_id="CAC09234.1"				
	/db_xref="GI:10189811"				
	/translation="MMKLLILIAAFAVYASADPIHYDKITEINKAIDEAVAAIERSE				
	TFDPKVPDHSDFEIRHIGLIDLGQDMNIIQVGLKQMKRGVDANVKSDESVYKH				
	LFVGHDDVVMSEYEDLAVKLGDLIPNTHVLSIDDFEVLSEYSEEGNMTLSFEVR				
	OFANVNVHIGSLSLDPIFNVLSVLTAFIOPDYRAMTVLNPAPFKELERNNO"				
	119				
BASE COUNT	270 a 133 c 142 g 267 t				
ORIGIN					
	mat.peptide				
	119				
Alignment Scores:					
Pred. No.:	2.78e-74	length:	812		
Score:	959.00	Matches:	182		
Percent Similarity:	94.37%	Conservative:	19		
Best Local Similarity:	85.45%	Mismatches:	12		
Query Match:	89.79%	Indels:	0		
DB:	6	Gaps:	0		
US-10-024-955-7 (1-213) x AX028697 (1-812)					
QY	1 MetMetLysPheLeuLeuIleAlaIleAlaValAlaPheValAlaValSerAlaSPProIle	20			
DB	68 ATGATGAATTTATTTATTTGATTTCTGCCGACGTTTGTTCGGCTTGGCGATCCCAATT	127			
QY	21 HlSTYrASpLysIleThGluGluIleasnLysAlaIleaspAspAlaIleAlaIle	40			
DB	128 CACTATGATAAATCAACGAGAACAAATTAACAAAGCTGTGATGAAGCCGTCGCAATT	187			
QY	41 GllugInserGluThrIleaspProMetLysValProAspHisAlaSPlysPheGluArg	60			
DB	186 GAAAAATCCGAACACTTGCATCAATGAAGGTACCCGATCATTCGATTAATTCGAACGA	247			
QY	61 HisValIleValIleValAspPheLysGlyLuleuAlaMetArgAsnIleGluAlaArgGly	80			
DB	248 CATATTGGATCATCGATTTAAGAAGTCATTAAGACATTCGCAACATTCAGTTCGACGA	307			
QY	81 LeuLysGlnMetLysArgGlnGlyLysAlaAsnValLysGlyGluGluGlyIleValLys	100			
DB	308 TTTAAACAATGAAACGTCGTAGCTATCTAATGTGAAAGAGGAAGATGGCTGTCAA	367			
QY	101 AlaHISLeuLeuIleGlyValAlaHisaspIleValISerMetGluTYrAspLeuAlaTYr	120			
DB	368 GCTCATTTGTTGGTGGGGTTCATATGACGCTGTTCATATGGAATTAATTAAGCAATC	427			
QY	121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValAla	140			
DB	428 AATATTGGGATCTTCATCCAAACATGATGATTCGATATTCAGAGATTTTGTTGTC	487			
QY	141 AlaLeuSerLeuGluIleSerAspGluLysAsnIleThrMetThrSerPheGluValArg	160			

Db 488 GAATTATCGCTGAGTAGGACGAGGTAATATGACATGACATCGTTGCAAGTACGT 547
 Qy 161 GlnpheaLaasnValValaLnsHsIsleGlyLeuSerIleLeuAspProIlePheGly 180
 Db 548 CAATTGGCAATGTTGTCATCATATGTTGGTCTTTCAATTTTGCATTTTCGCT 607
 Qy 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys 200
 Db 608 GTCTATCCGATGTTTGGACCGCTATTTTCCAGGATACCGTACGTGACAAATGACCAAA 667
 Qy 201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
 Db 668 GTATTGGCACCAGCATTCAAAAAGAAATGGAACGAAAC 706
 RESULT 6
 LOCUS LDE271058 767 bp mRNA linear INV 09-FEB-2001
 DEFINITION Lepidoglyphus destructor mRNA for putative allergen (belongs to group 7 allergens).
 ACCESSION AJ271058
 VERSION AJ271058.1 GI:6706281
 KEYWORDS allergen; group 7 allergens.
 SOURCE Lepidoglyphus destructor.
 ORGANISM Lepidoglyphus destructor.
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Acariformes; Sarcophormes; Astigmata; Glyciphagoides; Glyciphagidae; Lepidoglyphus.
 REFERENCE 1 (bases 1 to 767)
 AUTHORS Eriksson,T.L., Rasool,O., Huecas,S., Whitley,P., Cramer,I.R., Appenzeller,O., Gafvelin,G. and van Hage-Hamsten,M.
 TITLE Cloning of three new allergens from the dust mite Lepidoglyphus destructor using phase surface display technology
 JOURNAL Eur. J. Biochem. 268 (2), 287-294 (2001)
 MEDLINE 21099292
 PUBMED 11168362
 REFERENCE 2 (bases 1 to 767)
 AUTHORS Rasool,O.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2000) Rasool O., Unit of Clinical Allergy Research, Institution of Medicine, Clinical Allergy Research L2:04, Karolinska Hospital, S-171 76 Stockholm, SWEDEN
 FEATURES
 source
 Location/Qualifiers
 1..767
 /organism="Lepidoglyphus destructor"
 /db_xref="taxon:36936"
 /country="Sweden"
 54..704
 /note="belongs to group 7 allergens"
 /codon_start=1
 /product="putative allergen"
 /protein_id="CA65963.1"
 /db_xref="GI:6706282"
 /db_xref="SPTREMBL:Q9U1G2"
 /translation="MQLAIVAVIAGLGAASAAKPAYVDNMANQYDIKVSITTK KELDPEKTEQKVPIDKKIGLHIGKSAITINAVYITLSISRGDAKIDPDGGAFAA TRLGDKNIRIKTDLHLDKIHLNPKKEEDIDIMKLTSLKIDAEKPSLDDEID EPEQVELFIRHGLDPLVDYVIAQSPVKYFNPQARKLYTDMLKPLVEIKKLIN"
 54..110
 sig_peptide
 111..701
 mat_peptide
 /product="putative allergen"
 /note="belongs to group 7 allergens"
 BASE COUNT 259 a 172 c 149 g 187 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,89e-13 Length: 767
 Score: 251.50 Matches: 63
 Percent Similarity: 48.85% Conservative: 43
 Best Local Similarity: 29.03% Mismatches: 102
 Query Match: 23.55% Indels: 9
 Db: 3 Gaps: 5
 US-10-024-955-7 (1-213) x LDE271058 (1-767)

Qy 1 MetcLysPheLeuLeuIleAlaAla---ValAlaPheValAlaValSerAla----- 17
 Db 51 ATAAATGCAATATCTTCCATCTTCTGATGTTGGCCCTTGGCCGCTCATGCTCGCCG 110
 Qy 18 ---AspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAla 36
 Db 111 CATTAAGCCAGCCCTACTAGCAT-----GACAACATGGCCACACAGATGGTGACCATC 164
 Qy 37 IleAlaAlaIleGluGlnSerIleuThrIleAspPrometLysValProAspHisAlaAsp 56
 Db 165 GTCAAAAGTTTACCACCAAAAGAACACGATCGCTCCATGATGACAAACCAAGTT 224
 Qy 57 LysPheGluArgHisValGlyIleValAspPheLysGlyLeuAlaMetArgAsnIle 76
 Db 225 CCAATTGACAAAAGATTGGATGATCCATATCAAAAGTAGTGCACCACTTAAAGATGCC 284
 Qy 77 GluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGlu 96
 Db 285 GTGATTTACCGGCTTACGCATATCTACAGACCGGATGATGCCAAGATGCACACCGCGT 344
 Qy 97 GlyIleValLysAlaHisLeuLeuIleGlyValHisAspPheIleValSerMetGluTyr 116
 Db 345 GCGCGCTTGTCTGCAACCCCTTAAATTGGCC-----GACAAAAGCATTCGAATCAAAACC 398
 Qy 117 AspleuAlaTyrLysLeuGlyAsp---LeuHisProThrThrHisValIleSerAspIle 135
 Db 399 GATCTTCATCTGACCCCTTGGCAAAATTTATCTATCCCAATCTTAACTGCAGGCACACATT 458
 Qy 136 GlnAspPheValAlaAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThr 155
 Db 459 GCGCATATGACATGAAATGAAATTTGAAGTTGACAGCCGAGGACGACGATCGTTGGAC 518
 Qy 156 SerPheGluValArgGlnPheAlaAsnValValaHisIleGlyLeuSerIleLeu 175
 Db 519 CAATTGAGATTGACGAATTCGACAGGTCGACTGTTATCCATGCTGGGACCGTTG 578
 Qy 176 AspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArg 195
 Db 579 GACCCGCTGGTGACGATGTCGATTCATTTGTCAGATATTTCAATTCACACAGCCCGCA 638
 Qy 196 LysGluMetThrLysValLeuAlaProAlaPheLysArgGluLeuGluLys 212
 Db 639 AAATTGGTCACCGACATGTTGAAGCCAAATTCGTGTCGAAGACATCAAAAAA 689
 RESULT 7
 ATFF8M21
 LOCUS Arabidopsis thaliana DNA chromosome 5, BAC clone F8M21 (ESSA project).
 DEFINITION
 ACCESSION AL353993
 VERSION AL353993.1 GI:7671480
 KEYWORDS
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 103996)
 AUTHORS Bevan,M., Murphy,G., Ridley,P., Hudson,S., Bancroft,I., Mewes,H.W., Rudd,S., Lemke,K. and Mayer,K.F.X.
 JOURNAL unpublshed
 JOURNAL 2 (bases 1 to 103996)
 REFERENCE Arabidopsis sequencing project.
 AUTHORS Direct Submission
 TITLE Submitted (28-APR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail: lemke@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7JY Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk
 COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4

FEATURES and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

source Location/Qualifiers

1. .103996
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
2726. .4263
/gene="F8M21_10"
join(2726. .2996,3080. .3207,3291. .3420,3829. .4050,4133. .4263)
/gene="F8M21_10"
/note="similarity to predicted proteins, Arabidopsis thaliana and Drosophila melanogaster"
/codon_start=1
/product="putative protein"
/protein_id="CAB89322.1"
/db_xref="GI:7671481"
/translation="MGFEKPEKEVLEILSKNCKSNFNSVKKKKKKKKMMATWR
RKRIISPADGITAVRRLENTCKEVPFNSGGPGVIPSDEKIQQLREILDDKKPEDVGLTP
MPYFERNNGVEARSSPITVYLHLHQCDFSGVIGFCLPSPGVIPLHHHGMVFSKIL
FGTMHKSVDYVADAMPDSTRLAKLYDSFTFACNASTILYEDGGMHRTATTA
CAVDLYGPPYCNPEGRCOTYFLEPFLDKLSSPDDVLSSEKEGYAWLQGERDNP
DHRTNYGALYRCRYED"
2726. .2996
/gene="F8M21_10"
/number=1
2997. .3079
/gene="F8M21_10"
/number=1
3080. .3207
/gene="F8M21_10"
/number=2
3208. .3290
/gene="F8M21_10"
/number=2
3291. .3420
/gene="F8M21_10"
/number=3
3421. .3828
/gene="F8M21_10"
/number=3
3829. .4050
/gene="F8M21_10"
/number=4
4051. .4132
/gene="F8M21_10"
/number=4
4133. .4263
/gene="F8M21_10"
/number=5
8338. .10791
/gene="F8M21_20"
join(8338. .8421,8816. .9406,9522. .9635,9934. .10791)
/gene="F8M21_20"
/note="similarity to TMV response-related gene product, Nicotiana tabacum, EMBL:AB024510"
/codon_start=1
/product="putative protein"
/protein_id="CAB89323.1"
/db_xref="GI:7671482"
/translation="MEVLKLPSSSEPLKDKFGSVQIHEANKGDGDHOLESAKAMS
EYKENELKGLMERIESDYKSLKRFEDIIQEPNTATKQNMVDHKKPTTDDLS
FPOERELVSLIGRSSPSDPVKKEETDAISAVNDEELTKAGLILGINGVNG
EPKEGLSMENRANSSSEAMAPGVKGRSPAPASGDADGAGQONNIVARRVVR
ARCDPTMNDGCMRKYGOKIAKGNCPRAIYRCIYAPGCPVRKQYORADMSILIT
TYEGHSHSLPLSATTMASITSAASMLSGSSSSPAAEMIGNLIDNRFNNNSST
YSPTLHSPHPTVLDLTAPOHSSSSSSLSLNFKFSNSFQRPSTSLNFSSTST
SSNPSTLNLPAIMGNGSSYTPYNNVOGTSNLGKTVQNSQSLFETLTKALSDPS
FHSVIAAISTWNGSGEOIIVGRHSISNNIQOQTNNNNKCGCGYFSSLLMSNIMA
SNOTGASLDOPSSQLPFSMFKNSSSSSTTNFVNKEKS"
8338. .8421
/gene="F8M21_20"

intron /number=1
8422. .8815
/gene="F8M21_20"
/number=1
8816. .9406
/gene="F8M21_20"
/number=2
9407. .9521
/gene="F8M21_20"
/number=2
9522. .9635
/gene="F8M21_20"
/number=3
9636. .9933
/gene="F8M21_20"
/number=3
9934. .10791
/gene="F8M21_20"
/number=4
12598. .14582
/gene="F8M21_30"
join(12598. .13168,13276. .13325,13441. .13564,13657. .13764,13873. .14182,14273. .14582)
/gene="F8M21_30"
/note="similarity to probable aldose 1-epimerase, Nicotiana tabacum, PIR:T01933"
/codon_start=1
/product="putative aldose 1-epimerase-like protein"
/protein_id="CAB89324.1"
/db_xref="GI:7671483"
/translation="MARSSVLLCLILVTLGVIVSYTRADNVELSKSDLESSKKRV
DHDSTHKEKGGEVAGASDDEDNKKERKKEHDVQKKQHEKKDKDEKKNVKKK
SGCHDKDDDEDKKHKDKKKGHNDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
GLYELKGNLTWKFTNMGASISILHFPKNGKMDIIVLGDSVTKYKTDKVFYGATVG
RYANRIGKGFKLNGKEYKTSVNDGKNTLHGKKGGDVVAVAKHQYDKKPHIVFT
HSPDGDGFPGEISVTYTKLVKNELSVWEAKPKOKATPVNLAAHSHYMLGGHNS
GILSBEIOLISGTRPVDELIPTRKINPYGTAIDFLQLPPIKDNKDLKTGIDIN
YCLIDGAKKMRKIVELVDKRSGRKMKELSGNQGLDYYTGGMKLDVKKGKGAAYQAFGG
LCLEIQTSPDALNHPRFSPQIVPEGRKKYKHTMLFRFSIVS"
12598. .13168
/gene="F8M21_30"
/number=1
13169. .13275
/gene="F8M21_30"
/number=1
13276. .13325
/gene="F8M21_30"
/number=2
13326. .13440
/gene="F8M21_30"
/number=2
13441. .13564
/gene="F8M21_30"
/number=3
13565. .13656
/gene="F8M21_30"
/number=3
13657. .13764
/gene="F8M21_30"
/number=4
13765. .13872
/gene="F8M21_30"
/number=4
13873. .14182
/gene="F8M21_30"
/number=5
14183. .14272
/gene="F8M21_30"
/number=5
14273. .14582
/gene="F8M21_30"
/number=6
17863. .19172

gene

gene
CDS
/gene="F8M21_40"
complement(join(17863..18198,18753..19172))
/gene="F8M21_40"
complement(join(17863..18198,18753..19172))
/gene="F8M21_40"
/note="this sequence contains three differences at the amino acid level when compared to EMBL:AC008261. This may represent strain specific differences or sequencing errors."
/codon_start=1
/product="homebox-ziprin protein HA7"
/protein_id="CAB89325.1"
/db_xref="GI:7671484"
/translation="MNNRSMSFTGVSDDHNLTKSPPTTNNMNDODVGEEDINLSDGQSHMIAEGKKRIKRLNEQVRALEKSEFGKLERPKMOLAKALGIQROIALTPWPNRRARPKRTQQLERDIDLKQFDVLYKSDSDLSLANNKKIHAELVALKHKDKESAKIRREPEREASWNNSTENNNNNNSDANHVMISIKDLPPSSIRSATATTTSTHIDHQIVQDODGQFCNNMGNSIDETTSASVWAMPDQDQDQDQNNHNNHOFN"
complement(17863..18198)
/gene="F8M21_40"
/number=1
complement(18199..18752)
/gene="F8M21_40"
/number=1
complement(18753..19172)
/gene="F8M21_40"
/number=2
25399..26465

Alignment Scores:		
Pred. NO.:	497	Length: 103996
Score:	104.50	Matches: 39
Percent Similarity:	49.65%	Conservative: 52
Best Local Similarity:	27.27%	Mismatches: 33
Query Match:	9.78%	Indels: 19
DB:	8	Gaps: 6

US-10-024-955-7 (1-213) x ATF8M21 (1-103996)

QY 52 ProAspHisAlaAspLysPheGluArgHisValGlyTleValAlaAspPheLysGlyGluLeu 71

Db 63917 CCAAAATGCAGCGACACAAGGTGGAAGAGATTCTGACTTGTTCATTCCACCACCTCGTCTT 63

```

02  /2 AlaMetArgAsn-----IleGluAlaArgGly-----LeuLysGln 83
      ... .. ||| ... ..

```

DD 033/1 C1G1GCGAGG11C1AGG11G611G1A1CC11G611AAGGAGGAGCC1C1A1A1C1GAG 64

[illegible][illegible][illegible]

D6 64157 TTGCGAGTTTCTCCCTTCACAGCTTAAGGTACTTCAACAATGGAGCAGCTTAAACAGCAATTGA

QY 138 pheValValAlaLeuSer-----LeuGluIleSerAspGluGlyAsnIleThrMetThr 15

Db 64217 TTTGATGTT CAGATT TCTGGAGACTTGACGTTGCCAAGGAGCTCTCATACAATTACA 64

QY 156 SerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeu 177

Db 64277 TCA-----AGATTAAAGAGCCCAATGTTTTCGACCGAGAGGTCCTGTTCTCTGCACTT 64

QY	1/6 Asprolone 1/8
	:.

04320 AIGCCGAL 04330

AB016077

LOCUS	AB016077	3737 bp	DNA	linear	BCT 09-JUN-2001
DEFINITION	Streptococcus mutans genes for phosphoglycerate dehydrogenase, sakacin A production response regulator, complete cds.				
ACCESSION	AB016077				
VERSION	AB016077.1	GI:6681649			
KEYWORDS	sakacin A production response regulator; phosphoglycerate dehydrogenase.				
SOURCE	Streptococcus mutans (strain:MF8148) DNA, clone:PY570.				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
REFERENCE	1 (sites)				
AUTHORS	Kawabata,S., Terao,Y. and Hamada,S.				
TITLE	Molecular cloning, sequence and characterization of a novel streptococcal phosphoglycerate dehydrogenase gene				
JOURNAL	Oral Microbiol. Immunol. 15 (1), 58-62 (2000)				
MEDLINE	21069259				
REFERENCE	2 (bases 1 to 3737)				
AUTHORS	Terao,Y., Kawabata,S. and Hamada,S.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-JUL-1998) Yutaka Terao, Osaka University, Faculty of Dentistry, Department of Oral Microbiology, Yamadaoka 1-8, Suita, Osaka 565-0871, Japan (E-mail:tteraodent.osaka-u.ac.jp, Tel:06-879-2897)				
FEATURES	location/qualifiers				
source	1..3737				
	/organism="Streptococcus mutans"				
	/strain="MF8148"				
	/db_xref="taxon:1309"				
	/clone="PY570"				
	208..1518				
gene	/gene="pgda"				
CDS	208..1518				
	/gene="pgda"				
	/codon_start=1				
	/transl_table=1				
	/product="phosphoglycerate dehydrogenase"				
	/protein_id="BAA8823.1"				
	/db_xref="GI:6681650"				
	/translation="MALPVAIYGRPWGKSALPNRIAGRISIVEDVEGTPDRIRY KAERLNROFSIIDTGGIDVDAPPEMDIKQADINATPEADYIVFVVSAAKGTIDADEY VAILIRTRHPVILAVNKNVDNPEKRSATYDFVALGDPIVSSAHGIGTGDVLDIAIV DNLPAAEBOESSDIKPSLIGRPWGSLSLINALGEDRVIASVAGTVDADIDTFT DEDEQETIMIDTAGMRSGKGYENTEREYSVARAMRAIDRSDIYLVMLNABEGIREYDK RIGFAHEAGKGIYVYVNNKMDAIKKONRYAOMETDIRDNROYIIPYAIYVSATVTKO RIHKTPVIOYSGSONTRIPSSUNYVNDVAINTPPDCKRLKIFATQVSKP PPTVTFNEBELMHFTLRFLENDIROYAFVEGTPITLAKRR"				
gene	1675..2124				
	/gene="sapr"				
CDS	1675..2124				
	/gene="sapr"				
	/codon_start=1				
	/transl_table=1				
	/product="sakacin A production response regulator"				
	/protein_id="BAA8824.1"				
	/db_xref="GI:6681651"				
	/translation="MIALKTLGGSARAFETIARKVLAKVSPFEMRVSMMANAOAVY RHPHWCYRKQPEDNQDVGCLAIRIGNSANRPGSVESYTERKKSKATLAKQKIVLDI PLEPLTFYAOESSESHRVSGETEYRQMLKQKVAADGVRKGSFCSWP"				
BASE COUNT	1151 a	604 c	771 g	1211 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	12.5	Length:	3737		
Score:	103.00	Matches:	50		
Percent Similarity:	39.63%	Conservative:	36		
Best local Similarity:	23.04%	Mismatches:	77		
Query Match:	9.64%	Indels:	54		
DB:	1	Gaps:	11		
OS-10-024-955-7 (1-213) x AB016077 (1-3737)					
30 Asnysaiaiaieaspapalaiaiaiaiaiaieaguginsiergiutnriliaespromet 49					

```

Db      268 AATCGATATAGCTGTCGATTCATTCCTATTCCTGAGATGCTTGAAGGACGTAACAGAGAT 327
      50 LysValProaspHisAlaaspLysPheGluArgHisValGlyLeuValAspPheLysGly 69
      328 CGCATTTATACAAAGCGGAGATGCTTAATCGGCATTTTCATCATTCATGATACGGCGGT 367
Qy      70 GluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAsp 89
      388 -----ATTGACGATGTTGATGCCCTTATGACAAATTAAGCATCAAGCTGAT 438
Qy      90 AlaAsn-----ValLysGlyGluGlyIleVal 99
Db      439 ATTGCTATGACAGAGCGCTGATGTTGTTTGGTATCTGCTAAGGAGGATTCACA 498
Qy      100 Lys-----AlaHisLeuLeuIleGlyValHisAspIleVal----- 112
Db      499 GATGCCGATGATATGATGACAAAGATTCCTTATCGACACACAAGCGGTATTTAGCT 558
Qy      113 -----SerMetGluTyrAspLeu--AlaTyrLys 121
Db      559 GTTAATAAAGTTGATATCTCGATGCGCAGATGCTATTATGATTTCTATGCGCTTGGA 618
Qy      122 LeuGlyAspLeuHisPro-----ThrHisValIle-----SerAspIleGlnAsp 137
Db      619 CTGGGGGATCTTACCCCTGTTCTTCGTCATGATGTTGTCACAGAGATGCTTGTAT 678
Qy      138 PheValVal--AlaLeuSerLeuGluIleSerAspGlnLysAsnIleThrMetThrSer 156
Db      679 GCATGCTGATATATTTTACACAGAGCGCCAGAAAGAG-----AGT 720
Qy      157 PheGluValArgGlnPhe-----AlaAsnValVal 166
Db      721 TCAGACATTTATTAAGTTTATGTTGATCGACGCCCAATGTTGTAATTCAGTTTATC 780
Qy      167 AsnHisIleGlyLysLeuSer-----IleLeuAspProIlePheGlyValIleSerAsp 184
Db      781 AATGCTATCTTACGAGAGAATGCTGTTATGATCATCACCACTACTGCTAGACCCCGTAT 840
Qy      185 ValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLysVal 201
Db      841 GCGATTGATACACCTTCACGATGAAGAAGCCACAGAAATTTCACATGATT 891

RESULT 9
LOCUS      AE005025      11862 bp      DNA      linear      BCT 12-FEB-2001
DEFINITION Halobacterium sp. NRC-1 section 56 of 170 of the complete genome.
ACCESSION  AE005025 AE004437
VERSION     AE005025.1 GI:10580404
KEYWORDS
SOURCE      Halobacterium sp. NRC-1.
            Halobacterium sp. NRC-1.
            Archaea: Euryarchaeota: Halobacteriia: Halobacteriales:
            Halobacteriaceae: Halobacterium.
REFERENCE   1 (bases 1 to 11862)
            Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Sprogna,J.,
            Swartzell,S., Welt,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leitbauer,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            Dassarma,S.
            Genome sequence of Halobacterium species NRC-1
            Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
TITLE      Journal
PUBMED     20504483
AUTHORS     2 (bases 1 to 11862)
            Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Sprogna,J.,
            Swartzell,S., Welt,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leitbauer,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            Dassarma,S.
FEATURES
            source
            title      Direct Submission
            journal    Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            source     Roosevelt Way NE, Seattle, WA 98105, USA
            location   Location/Qualifiers
            1..11862
            /organism="Halobacterium sp. NRC-1"
            /strain="NRC-1"
            /db_xref="taxon:64091"
            67..957
            /gene="VNG0834C"
            67..957
            /gene="VNG0834C"
            /note="conserved hypothetical protein"
            /codon_start=1
            /transl_table=11
            /product="Vng0834C"
            /protein_id="AAG19288.1"
            /db_xref="GI:10580405"
            /translation="MPDTQLASLLOTAMKREASAVLESARTEGDIQMDVSDAVSS
            SOWGRILISAGVLTSTDTGFVADPDLAELEAHTDAESTETHTDALESASWTLYDK
            AAVAGVIGLIFAGYWNSSIOPTTASVDNOLAVMNLPHIVLVLAVTGYSTYLO
            SRLMDETLOAQKDRMAALKERREAKERGGDEALEERIQSEQMDAAGDGLGFKLOF
            PMWYIMILPIVPELIMRMEYRGHIGASHGWVPLAGIVANQHALIGMAWYIWMYF
            VCSMASRPLQKTFNIOAST"
            complement(969..1331)
            /gene="ldr2"
            /note="VNG0835G"
            complement(969..1331)
            /gene="ldr2"
            /note="VNG0835G"
            complement(969..1331)
            /gene="ldr2"
            /note="ldr2"
            /codon_start=1
            /transl_table=11
            /product="Iron-dependent repressor homolog"
            /protein_id="AAG19289.1"
            /db_xref="GI:10580406"
            /translation="MSGADVLYLTYIESHRDSDPVPNGQLADRLKAPSTYEMIKO
            LDENGLVYTTTPYKGRALTDAGKERATLHESYVYTSWFFRSVLDLTFHESEMEIASI
            IDPIYTDRLAATLPDDR"
            3036..3416
            /gene="VNG0836H"
            3036..3416
            /gene="VNG0836H"
            /note="hypothetical protein"
            /codon_start=1
            /transl_table=11
            /product="Vng0836H"
            /protein_id="AAG19290.1"
            /db_xref="GI:10580407"
            /translation="MKNLRCAVDVGREHRLAESHRLAIVLLIGCTTALADCVIRVF
            PQRDRTKERREGEYKELIMLPYFCDVLSNVVKKILIRCVDSPREHRGKLHPEYE
            EHRRODDPAVCYTRSLRPTSETA"
            4278..4568
            /gene="VNG0837H"
            4278..4568
            /gene="VNG0837H"
            /note="hypothetical protein"
            /codon_start=1
            /transl_table=11
            /product="Vng0837H"
            /protein_id="AAG19291.1"
            /db_xref="GI:10580408"
            /translation="MTQSEFLRTWVQAGRGLFSDSGSSDATPRGDALDRVQSM
            VDEGAVSWDELVAQLAGDFEDRIDTAVQSLADGVIRDPGRGVVSGADE"
            4561..5634
            /gene="ssrA"
            4561..5634
            /gene="ssrA"

```


gene
/note="Sra"
/codon_start=1
/transl_table=11
/product="Integrase/recombinase"
/protein_id="AAG19292.1"
/db_xref="GI:10580409"
/translation="MSPPRDSLARETDADADDPIGIFLEDMOYHGAKTRDAIYARY
LNSFAPLSNAGQAPVDAGRDCLTWHHLREQYAEStVATYARFYSYMOQGV
VGNPMALVSOEMDEADIDTNPTRDI SYAAMREPIADITHLERAVLVGLIKTLRNG
ELCANDLRDNLNDAIPASVSAEYGVGRLGSHPTLLVYDPMSEGETVNGOSASAK
RRARLTVPDLREKTELRYRLWLAIRDDTSPADPFVSGDAMGKRLPENVRGVRRH
AEPRWYKLGGAENATVPYHFRHFTTHLNRIGDGRVYKYLNGVADDIIDITYTHN
WDRKREYVLTNLSLTTHGGG"
complement(5898. .6002)
/gene="VNG0840H"
complement(5898. .6002)
/gene="VNG0840H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng0840h"
/protein_id="AAG19293.1"
/db_xref="GI:10580410"
/translation="MKNADILGNSDIYGLGGSALRKTLKKIDFTI"
complement(6277. .6939)
/gene="icfa"
/note="icfa"
complement(6277. .6939)
/gene="icfa"
/note="icfa"
/codon_start=1
/transl_table=11
/product="carbonic anhydrase"
/protein_id="AAG19294.1"
/db_xref="GI:10580411"
/translation="MLAGNQKHAEFQSRFDYVONAOHPDAVYTCGDSRYLDLMA
NTPDSIFTCANIGNRVIOQTPDGVVSDVLYPIEHGTEFAVVCNCGAATYAY
RNLDTDAEAPPGVIAHSLTPIRTRLAAGVDLLPDLSSAINLVLYVNDQVOFL
ESDRDDVVSVGAYVDFQDYPGSGRGAHVINVDGETPDLREQAPELASRITRLM
SY"
complement(7041. .7850)
/gene="VNG0845C"
complement(7041. .7850)
/gene="VNG0845C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng0845c"
/protein_id="AAG19295.1"
/db_xref="GI:10580412"
/translation="MNTAPDGTERTATYTSAGRPSLAPNLLPAIRLVASAGVDPG
NIVLDGGTGGAALTARRAGATYGLDVTGSMLELRADTQLTDVNDISWAGDGL
LPFPDADPVILSFGHYFSRAAEAGREMLRVAPGGRATFTAKSPDGLVGLTDL
DDYVADHDHPAHLRMGNPEFVRDHLGDAELTFERRIAFRVSEPHNRNFAEBA
GLSALPALADPGARSALREBALATLIDEMFADNTVREYLLARGVVDGA"
7986. .9461
/gene="VNG0846C"
7986. .9461
/gene="VNG0846C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng0846c"
/protein_id="AAG19296.1"
/db_xref="GI:10580413"
/translation="MRRFAVAASRLGVDTRSLAARVGLAVVYLVDLALRISQTAAY
TDSGLVPRVMAQOYPPVRSASLHMLSCAAMAQALRAVTAVALVIGYRSAAEF
ATGVLVSMHMANPLVNLGCDLVLOLFEVMAIGLPLREMSVDGAGANDRVRVSA
TGVGLCAIYITTSNAVLKLRGDANVSGTAIVRTVSELPFVGLGHSPPRLVAF
DRVYLVLVLAGSVLITLGMKRRAAAAAFACMHGMAQTQGLFPLVAYVGLVPPVP
GCWEAAAAARQRTVHFVRMTADDERAPSVPPSLDRSVARAKRTAVGVFVVLV
GQAATGVSGVETPVSVGPIADGADAPPWMYFAEPILSVDHVMAKATTPGGGRI
GAFAGATGSMGPPPVSAVPSARVRKYVVIADSDSRVAGLVAATLCRSSAAGTGH

gene	CDS
gene	CDS
<p>RRATNVTVYDVROPTRLGSGSEPTHRRIIGAAADCPVNAAEKV"</p> <p>complement(9486..10391)</p> <p>/gene="VNG0847h"</p> <p>complement(9486..10391)</p> <p>/gene="VNG0847h"</p> <p>/note="hypothetical protein"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="Vng0847h"</p> <p>/protein_id="AA019297.1"</p> <p>/db_xref="GI:10580414"</p> <p>/translation="MPRPVRIAYPRKGRPLEAVLDVNGELADAAVSADDITLRLRE</p> <p>KGLGKHHDHARGVYERLLAAVSPPDEPTADYTLRDSRGKPRRIYFDSITLVAGT</p> <p>RVELIGREEPRLARTHEHFAFGSDALVLEEVVGLAPPELTISADVNEIIDDVDV</p> <p>RVAAGLSDITYHTLLATPDHAGSGADRALVDLAYSQVCIISPERLVEAVLSTAYTE</p> <p>GLTPTPDADHDEEEAIDAGVAGCVLTIVGSPARHDGLTVGQLEPSEVLEENQHE</p> <p>ATGRVAVSRQALTTAPETFAIPQ"</p> <p>complement(10440..11789)</p> <p>/gene="VNG0849C"</p> <p>complement(10440..11789)</p> <p>/gene="VNG0849C"</p> <p>/note="conserved hypothetical protein"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="Vng0849c"</p> <p>/protein_id="AA019298.1"</p>	<p>gene</p> <p>CDS</p>

Alignment Scores:	
Pred. No.:	181
Score:	96.50
Percent Similarity:	37.18
Best Local Similarity:	20.51
Query Match:	9.04%
DB:	1
Length:	11862
Matches:	48
Conservative:	39
Mismatches:	100
Indels:	8
Gaps:	47

US-10-024-955-7 (1-213) x AE005025 (1-11862)

QY 13 ValAlaValSerAlaAspProIleHisTyrGrasPlsYsIleThnGluGluIuIleAsnLysAla 32
 11003 GTGGGCGGCTGCTGGGCGGAGTATCGGCAAGCAGAGCGCCGCGGCGGCGGAGGACCTCGGCAACC 10944
 Pk 11003 GTGGGCGGCTGCTGGGCGGAGTATCGGCAAGCAGAGCGCCGCGGCGGCGGAGGACCTCGGCAACC 10944

33 IleasppAlaIleAlaIleGluInserGluThrIleasppPrometIysValPro 52

Db	10943	ATCCAGGACGGCGATGGTGGGGTACCCGAAATGACGGCGGTGGAGCTCACACACCTGCG	10884
Ov	53	ASDHSIAASDIPVPHGILNRAHISVA]-----	62

Db 10883 TCCCGGCTCCGAGATCGAAGCGTACACCGAGCCCTGAGCTGTTTCATCGACGACAAAC 10824

Dy 63 -----glyllevalasphnelysglyglu 70
Db 10823 GGCACGCGACAGGACGTGCTCAACGACCCTCGAAGACGACGTGGCCGACCTGGAAACCGAG 10764

QY 71 LeuAlaMetArgAsn1Ieg1uAlaArgC1yLeuLysGln--MetLysArgGlnGlyAsp 89
||||: |||:|||||: |||

Db 10763 CTATCGACGGTGGGGCAGCAGCGTGGCGTGCAGCAGTCCGTCGACGCTGCCGCCGAC 10704

QY 90 AIAASnValLysGLIyGluGluGluIleValAlLysAlaHisIstLeuLeuIleGlyValHisasp 109

Db 10703 GACCGGGGCGGCGCTGCACGAGCGCGTCGACGACGTCGAAGCCACGCGGTGGGGCTGTGCAC 10644

Qy	110	AspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGluYAspLeuHisProThrThr	129
		:::	
Db	10643	GAC-----CTCGAATCGTCTGACCGGCTCGCGGCGACCGTGGACCGCCCTTCGAC	10593

```

QY      130 HisValIleSeraspIleGlnaspPheValAlaAlaLeuSerLeuGluIleLeuSeraspGlu 149
          ::::::::::: ::::: ||| ::::::::::: ::::::::::: :::::::::::

```

Qy	150	GlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIle	169
Db	10592	GAGCGCCCTGCGGACACCGAGGAGCGAGCGTGTCTGGCGCTGGCC	10542
		--GAGGTCTCC-----	

Db 10541 GCGGACGTGGAGCGCCCTCGAAGCGCCGACGTCTCGAGACACAGTCAGCGCATC 10482

```

QY 170 GLYGLYLeuSerIleLeuAspProIlePheGlyValLeuSerAspVal-----185
Db 10481 -----GACGACTGGCGCCACACAGCTGTCGAGCGTTTCAGCGCTGA 10440
QY 186 -----LeuThrAlaIlePhe-----GlnAspThrValArgLysGlu 197
Db 10439 GCGCTGGGATTCACACAGGTTTTCGGCGCCACTGGCTGCACCAACAAATGACACCCCG 10380
QY 198 MetThrLysValLeuAlaProAlaPheLysArgGluLeuGlu 211
Db 10379 CCGGTTGCAATCGCCGCCGGAAGACGGCATTTGAG 10338

RESULT 10
AX350829 892 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 29 from Patent WO0179294.
DEFINITION AX350829
ACCESSION AX350829
VERSION AX350829.1 GI:18616300
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1
AUTHORS
Taupier,R.J., Vernet,C.A., Fernandes,F., Shinkets,R.A.,
Majumder,K., Padigaru,M., Colman,S.D., Zehrusen,B.D., Spytek,K.A.,
Burgess,C.E. and Liu,X.
Novel human proteins, polynucleotides encoding them and methods of
using the same
Patent: WO 0179294-A 29 25-OCT-2001;
JOURNAL
Curagen Corporation (US)
FEATURES
source 1.892
/organism="unidentified"
/db_xref="taxon:32644"
/note="POLYX hmb0778p17_A1"
BASE COUNT 355 a 130 c 185 g 222 t
ORIGIN

Alignment Scores:
Pred. No.: 9.03 Length: 892
Score: 96.00 Matches: 44
Percent Similarity: 40.54% Conservative: 31
Best Local Similarity: 23.78% Mismatches: 58
Query Match: 8.99% Indels: 52
DB: Gaps: 10

US-10-024-955-7 (1-213) x AX350829 (1-892)
QY 27 GUGUuIleAsnLysAlaIleAspAspAlaIleAlaIleGluInserGluThrIle 46
Db 310 GAATACATCAACAGAGTTTGAATGATTGTAAGACAGTTAAAGTCAAGAGTTGAA 369
QY 47 Asp---PrometLysVal-----ProAspHisAlaAspLysPheGlu 59
Db 370 AATGGTCCATCTTCAGTGCACACAGATCTTAATCTCAGCATGCTGCAATGTTTC--- 426
QY 60 ArgHisValGlyIleValAspPhe-----LysGlyGlu--- 70
Db 427 CGGCATTTTCAGCAAAATCATGTTTATATACATGACACATAGACAGCAAGCAAGAAAG 486
QY 71 -----LeuAlaMetArgAsnIleGluAlaArgGly----- 80
Db 487 TGCAGACATTTTATTTCCTCAGCTTGAAGTTCCTGGAAGAGATGCTGAAGAGAT 546
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu----- 96
Db 547 GTAATGATATGCTTCATCAAGGAAATGGGAAAGTTTAAATGAAGCTTACTTACGAA 606
QY 97 ---GlyLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu----- 96
Db 607 ATCAATATCACTAAGACAACTTTTCAGAGATTTCACAGACACCAAGCAAGCAACTGTTAAT 666
QY 114 MetGluThrLysAspLeuAlaIleThrLysLeuGlyAspLeuHisProThrThrHisValIleSer 133

```

```

Db 667 TTGAG-----AACCAATTAAG 684
QY 134 AspIleGlnAspPheValAlaIleuSerLeuGluIleSerAspLysGlu----- 150
Db 685 GATTTAAGGATCTTTTCATTCAGATATCTCTTTAGTAGAGAGACAGAGACATC 744
QY 151 ---AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValAlaAsnHisIle 169
Db 745 AACAAATATGAATAGACAGATAGTACAAAGAGTATGTTAAACAATACTAAGAGAAA 804
QY 170 GLYGLYLeuSerIle 174
Db 805 TTTGACTACCTGTA 819

RESULT 11
AX350831 892 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 31 from Patent WO0179294.
DEFINITION AX350831
ACCESSION AX350831
VERSION AX350831.1 GI:18616301
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1
AUTHORS
Taupier,R.J., Vernet,C.A., Fernandes,F., Shinkets,R.A.,
Majumder,K., Padigaru,M., Colman,S.D., Zehrusen,B.D., Spytek,K.A.,
Burgess,C.E. and Liu,X.
Novel human proteins, polynucleotides encoding them and methods of
using the same
Patent: WO 0179294-A 31 25-OCT-2001;
JOURNAL
Curagen Corporation (US)
FEATURES
source 1.892
/organism="unidentified"
/db_xref="taxon:32644"
/note="POLYX CG55655_02"
BASE COUNT 355 a 130 c 185 g 222 t
ORIGIN

Alignment Scores:
Pred. No.: 9.03 Length: 892
Score: 96.00 Matches: 44
Percent Similarity: 40.54% Conservative: 31
Best Local Similarity: 23.78% Mismatches: 58
Query Match: 8.99% Indels: 52
DB: Gaps: 10

US-10-024-955-7 (1-213) x AX350831 (1-892)
QY 27 GUGUuIleAsnLysAlaIleAspAspAlaIleAlaIleGluInserGluThrIle 46
Db 310 GAATACATCAACAGAGTTTGAATGATTGTAAGACAGTTAAAGTCAAGAGTTGAA 369
QY 47 Asp---PrometLysVal-----ProAspHisAlaAspLysPheGlu 59
Db 370 AATGGTCCATCTTCAGTGCACACAGATCTTAATCTCAGCATGCTGCAATGTTTC--- 426
QY 60 ArgHisValGlyIleValAspPhe-----LysGlyGlu--- 70
Db 427 CGGCATTTTCAGCAAAATCATGTTTATATACATGACACATAGACAGCAAGCAAGAAAG 486
QY 71 -----LeuAlaMetArgAsnIleGluAlaArgGly----- 80
Db 487 TGCAGACATTTTATTTCCTCAGCTTGAAGTTCCTGGAAGAGATGCTGAAGAGAT 546
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu----- 96
Db 547 GTAATGATATGCTTCATCAAGGAAATGGGAAAGTTTAAATGAAGCTTACTTACGAA 606
QY 97 ---GlyLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu----- 96
Db 607 ATCAATATCACTAAGACAACTTTTCAGAGATTTCACAGACACCAAGCAAGCAACTGTTAAT 666
QY 114 MetGluThrLysAspLeuAlaIleThrLysLeuGlyAspLeuHisProThrThrHisValIleSer 133

```

```

Db 607 ATCAATATCACTAAGACCAACTTTTCAGAGATTGAACAGACAGACAGAACTGTAAAT 666
Oy 114 MetGUtYrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThHisValIleSer 133
Db 667 TTGGAG-----AACCAATAAAG 684
Oy 134 AspIleGlnAspPheValAlaLeuSerLeuGluIleSerAspGluGly----- 150
Db 685 GATTAAAGGATCTTTTCATTCAGATATCTCTTTAGTAGAGAAACAGACAGAGCATC 744
Oy 151 ---AsnIleThrMetThSerPheGluValArgGlnPheAlaAsnValValAsnHisIle 169
Db 745 AACCAATATTGAATGACGTAGTACAAAGAGTATGTTAACAAATACTAAAGAGAA 804
Oy 170 GlyGlyLeuSerIle 174
Db 805 TTGGACTAGCTGTA 819

RESULT 12
AX350827 893 bp DNA linear PAT 06-FEB-2002
LOCUS DEFINITION Sequence 27 from Patent WO0179294.
ACCESSION AX350827
VERSION AX350827.1 GI:18616299
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1
AUTHORS Taupier,R.J., Vernet,C.A., Fernandes,E., Shinkets,R.A.,
Majumder,K., Padiganu,M., Colman,S.D., Zerhusen,B.D., Spletke,K.A.,
Burgess,C.E. and Liu,X.
TITLE Novel human proteins, polynucleotides encoding them and methods of
using the same
JOURNAL Patent: WO 0179294-A 27 25-OCT-2001;
Curagen Corporation (US)
FEATURES
source 1..893
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
/note="POLYX h_n0778p17_A"
BASE COUNT 355 a 130 c 186 g 222 t
ORIGIN
Alignment Scores:
Pred. No.: 9.05 Length: 893
Score: 96.00 Matches: 44
Percent Similarity: 40.54% Conservative: 31
Best Local Similarity: 23.78% Mismatches: 58
Query Match: 8.99% Indels: 52
DB: 6 Gaps: 10

US-10-024-955-7 (1-213) x AX350827 (1-893)
Oy 27 GluGluIleAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIle 46
Db 311 GAAATCAATCAACAGACAGAGTTTGATGATTTAGTTAAAGAGTTAAAGAGTCAGAGGTGTA 370
Oy 47 Asp---PromElyVal-----ProAspHisAlaAspLysPheGlu 59
Db 371 AATGTCATCTTCAGTGCAGTGCACAGATCTTAATCTCAGCATGCTGCATGTTTC--- 427
Oy 60 ArgHisValGlyIleValAspPhe-----LysGlyGlu--- 70
Db 428 CGCCATTTTCAGCAATCATGTTTATATACATGACACAAATGACAGCAAGAGAGAG 487
Oy 71 -----LeuAlaMetArgAsnIleGluAlaArgGly----- 80
Db 488 TGCAGACATTTATTATTTCGTCAGCTGAGAGTTGCTGGAAAAGAGATGCTGGAAGAGAT 547
Oy 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu----- 96
Db 548 GTAATATGATATGCTTCATCAACAGAAATGGGAAGTTTTTAAATGAAGCTTACTACAGA 607

```

```

Oy 97 ---GlyIleValLysAlaHisLeu-----LeuIleGlyValHisAspAspIleValSer 113
Db 608 ATCAATATCACTAAGACCAACTTTTCAGAGATTGAACAGACAGACAGAACTGTAAAT 667
Oy 114 MetGUtYrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThHisValIleSer 133
Db 668 TTGGAG-----AACCAATAAAG 685
Oy 134 AspIleGlnAspPheValAlaLeuSerLeuGluIleSerAspGluGly----- 150
Db 686 GATTAAAGGATCTTTTCATTCAGATATCTCTTTAGTAGAGAAACAGACAGAGCATC 745
Oy 151 ---AsnIleThrMetThSerPheGluValArgGlnPheAlaAsnValValAsnHisIle 169
Db 746 AACCAATATTGAATGACGTAGTACAAAGAGTATGTTAACAAATACTAAAGAGAA 805
Oy 170 GlyGlyLeuSerIle 174
Db 806 TTGGACTAGCTGTA 820

RESULT 13
BC034696 1198 bp mRNA linear PRI 16-AUG-2002
LOCUS DEFINITION Homo sapiens, Similar to RIKEN cDNA A030009B12 gene, clone
ACCESSION BC034696
VERSION BC034696.1 GI:21961210
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 1198)
JOURNAL Strausberg,R.
Direct Submission
Submitted (24-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnarlane, P.H., Garcia, A.M., Lu, X., Huylk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov
Series: IRAC Plate: 28 Row: m Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Genomescan gene prediction, similarity but not identity
to protein.
FEATURES
source 1..1198
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:21382 IMAGE:4746351"
/tissue_type="Skin, normal"
/clone_id="NCL_CGAP_Skn3"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6, ccdb"
258..1142
/codon_start=1
CDS

```

```

/product="Similar to RIKEN CDNA A030009B12 gene"
/protein_id="AAH34696.1"
/db_xref="GI:21961211"
/translation="MKDRLOELKORTKELEISRDHSVSTTEPEOGVFIQAVIERE
PYAEHHEIKLOESINNLANNVOKFOQOASLYASMRERELKRESTITREIKIOA
EYINSNDLVAEVKSEYENGSPSSVTRILKSQJAHMRHQIQTIFYNDITAKOE
KCKTFILROLEVAEKEMSEEDVNDMLHQQKWEFNSLITEINIKAOISEIOHKE
LYVLENOIKDLDFLOISLVEEOGESINNIEMTVNSTKEYVNTKKEKGLAVKXK
RMRVILCMCCPCSSK"
BASE COUNT      498 a      173 c      231 g      296 t
ORIGIN
Alignment Scores:
Pred. No.:      12.9      Length:      1198
Score:          96.00      Matches:      44
Percent Similarity: 40.54%      Conservative: 31
Best Local Similarity: 23.78%      Mismatches: 58
Query Match:      8.99%      Indels:      52
DB:                Gaps:      10
US-10-024-955-7 (1-213) x BC034696 (1-1198)
QY 27 GUGLUILLASNLysAlaIleAspAlaIleAlaIleGluGlnSerGluThrIle 46
Db 564 GAATACATCAACAGAAAGTTGAAATGATTAGTTAAAGAAAGTTAAAGAGGTTGAA 623
QY 47 Asp---PrometLysVal-----ProAspHisAlaAspLysPheGlu 59
Db 624 AATGTCACATCTTCAGTGCACAGAGATACTTAATCTCAGATGCTGCAATGTTTC 660
QY 60 ArgHisValGlyIleValAspHe-----LysGlyGlu--- 70
Db 681 CGCCATTTTCAGCAAAATCATGTTTATATACAAATGACACATAGCAGCAAGAGAGAG 740
QY 71 -----LeuAlaMetArgAsnIleGluAlaArgGly----- 80
Db 741 TGCAGACATTTAATTTTACGTCAGCTTGAAAGTTGCGGAAAGAGATGCTGAAGAGAT 800
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu----- 96
Db 801 GTAATGATATGCTTCATCAGAGAAATGGAGATTTTTATGAAAGCTTACTTACAGAA 860
QY 97 ---GlyIleValLysAlaHisLeu-----LeuIleGlyValHisAspAspIleValSer 113
Db 861 ATCAATATCTACTTAACACACACTTTCAGAGATTGAAACAGACAGCAAGAACTGTTAAT 920
QY 114 MetGluTyrAspLeuAlaIleTyrLysLeuGlyAspLeuHisProThrHisValIleSer 133
Db 921 TTGGAG-----AACCAAAATTAAG 938
QY 134 AspIleGlnAspPheValAlaAlaLeuSerLeuGluIleSerAspGluGly----- 150
Db 939 GATTTAAGGATCTTTTCATTCAGATATCTCTTTAGTAGAGACAGACAGAGAGAGATC 998
QY 151 ---AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValAlaAsnHisIle 169
Db 999 AATTAATAATGAATGACAGAGATAGTACAAAGAGATGATTAAACATACTAAAGAGAAA 1058
QY 170 GlyGlyLeuSerIle 174
Db 1059 TTGGAGCTAGCTGTA 1073
RESULT 14
AX430414      1233 bp      DNA      linear      PAT 28-JUN-2002
LOCUS
DEFINITION Sequence 196 from Patent WO0240715.
ACCESSION AX430414
VERSION AX430414.1 GI:21655778
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
TITLE Molacup, for disease detection and treatment
JOURNAL Patent: WO 0240715-A 196 23-MAY-2002;
INCYTE GENOMICS INC (US)
FEATURES
source 1. 1233
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: LI-204626.1:2000SEP08"
BASE COUNT      489 a      184 c      256 g      304 t
ORIGIN
Alignment Scores:
Pred. No.:      13.3      Length:      1233
Score:          96.00      Matches:      44
Percent Similarity: 40.54%      Conservative: 31
Best Local Similarity: 23.78%      Mismatches: 58
Query Match:      8.99%      Indels:      52
DB:                Gaps:      10
US-10-024-955-7 (1-213) x AX430414 (1-1233)
QY 27 GUGLUILLASNLysAlaIleAspAlaIleAlaIleGluGlnSerGluThrIle 46
Db 643 GAATACATCAACAGAAAGTTGAAATGATTAGTTAAAGAAAGTTAAAGAGGTTGAA 702
QY 47 Asp---PrometLysVal-----ProAspHisAlaAspLysPheGlu 59
Db 703 AATGTCACATCTTCAGTGCACAGAGATACTTAATCTCAGATGCTGCAATGTTTC 759
QY 60 ArgHisValGlyIleValAspHe-----LysGlyGlu--- 70
Db 760 CGCCATTTTCAGCAAAATCATGTTTATATACAAATGACACAAATAGCAGCAAGAGAGAG 819
QY 71 -----LeuAlaMetArgAsnIleGluAlaArgGly----- 80
Db 820 TGCAGACATTTAATTTTACGTCAGCTTGAAAGTTGCGGAAAGAGATGCTGAAGAGAT 879
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu----- 96
Db 880 GTAATGATATGCTTCATCAGAGAAATGGAGATTTTTAAATGAAGCTTACTTACAGAA 939
QY 97 ---GlyIleValLysAlaHisLeu-----LeuIleGlyValHisAspAspIleValSer 113
Db 940 ATCAATATCTACTTAACACACACTTTCAGAGATTGAAACAGACAGCAAGAACTGTTAAT 999
QY 114 MetGluTyrAspLeuAlaIleTyrLysLeuGlyAspLeuHisProThrHisValIleSer 133
Db 1000 TTGGAG-----AACCAAAATTAAG 1017
QY 134 AspIleGlnAspPheValAlaAlaLeuSerLeuGluIleSerAspGluGly----- 150
Db 1018 GATTTAAGGATCTTTTCATTCAGATATCTCTTTAGTAGAGACAGACAGAGAGAGATC 1077
QY 151 ---AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValAlaAsnHisIle 169
Db 1078 AATTAATAATGAATGACAGAGATAGTACAAAGAGATGATTAAACATACTAAAGAGAAA 1137
QY 170 GlyGlyLeuSerIle 174
Db 1138 TTGGAGCTAGCTGTA 1152
RESULT 15
AC026300      23269 bp      DNA      linear      HTG 16-JUN-2000
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-778P17, WORKING DRAFT
ACCESSION AC026300
VERSION AC026300.3 GI:8568964
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

```


CDS	/gene="cpn60"
	<1..>558
	/gene="cpn60"
	/note="Cpn60; hsp60; groEL"
	/codon_start=1
	/transl_table=11
	/product="60 kDa chaperonin"
	/protein_id="AA049348.1"
	/db_xref="GI:21359246"
	/translation="ATVLAQAIIIGVKLKNYAGANPMDLRKIDRAKAVVENIAEQAEVDKFEIKPIEVAKISMGDEALGILAEAMOVKKRGVTVEAGKTETTVVEGE
	MOPRGIYSPEYEMNSDKMEASLDSPVLITDRKISTPKDLPLIEPIARCKSLIITI
	AEDVDGERLTLLTYNRLRGITKYAAV"
BASE COUNT	145 a 108 c 165 g 140 t
ORIGIN	
Alignment Scores:	
Pred. No.:	5.69 Length: 558
Score:	95.50 Matches: 42
Percent Similarity:	41.81% Conservative: 32
Best Local Similarity:	23.73% Mismatches: 50
Query Match:	8.94% Indels: 53
DB:	Gaps: 10
US-10-024-955-7 (1-213) x AF437087 (1-558)	
OY	13 ValAlaValSerAlaAspProLeuHisThrGlyAspLysIleThrGlucLuIleasnLysAla 32
Dd	46 GTAACAGCAGCGGTCCGAATCCGATG-----GATTGGAAGCGTGGT 84
OY	33 IleAspAspAlaTleAlaAlaIle-----GlucLISerGluThrIleasp 47
Dd	85 ATCGCAAGGCCCGGTGCCTTAAGGCTTGAGAACAATCGCGTAGCACGCTGACGAGGTT-- 141
OY	48 PrometLysValProAspHisAlaSprLysPheGluArg-----HisValGlyIleVal 65
Dd	142 -----GGCACAAGATTCGAGAAGATCGAACAATCGAACATACCGAAGATC 180
OY	66 AspPheLysGlyLysIleuAlaMetArgasnIleGluAlaArgGlyLysIleuLysGlnMetLys 85
Dd	181 TCTGCCAATGGCGGAGGAGCGCATCGGTAAAGTTGATGATCGGTGACGCGTATGCAAG 240
OY	86 ArgGngIglyAspAlaAsnVal-----LysGlyGluGluGlyIleValLysAlaHis 102
Dd	241 AAAGAGGCTGTCATCATCTTGAGAGGCTGAAGCGTACTGACATACATACAGTGC----- 291
OY	103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyAspLeuAlaTrpLysLeu 122
Dd	292 -----GATGTCTGTGAGGCTATGCGATTCGACAGAGCTAT----- 327
OY	123 GlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValValAlaLeu 142
Dd	328 -----ATCTGCTTACTTTATGACTAATTTGAC----- 357
OY	143 SerLeuGluIleSer---AspGluGlyAsnIleThrMetThrSerPheGluValArgGln 161
Dd	358 AAGATGGAAGCTAGCGCTGCACAGCCATATGCTTTATTATACGACACAAGAATTTCTACT 417
OY	162 PheAlaAsnValValAsnHisIleLeuGlyLeuSerIleLeuAspProIle 178
Dd	418 ATGAAGAGATCTT-----CTTCTCATTCCTTCGACCGTATTC 450
RESULT 20	
LOCUS	MPULM03 315079 bp DNA linear BCT 14-MAY-2001
DEFINITION	Mycoplasma pulmonis (strain OAB CTIP) complete genome; segment 3/3.
VERSION	AL445565 AL445566
KEYWORDS	AL445565.1 GI:14089942
SOURCE	.
ORGANISM	Mycoplasma pulmonis.
REFERENCE	Bacteria; Filicutes; Mollicutes; Mycoplastmataceae; Mycoplasma.
	1 (bases 1 to 315079)

[illegible]

JOURNAL Submitted (28-AUG-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, rbluel@tigr.org
 On Jan 5, 2001 this sequence version replaced gi:12025620.
 Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0013M12 is from Oryza sativa chromosome 3
 The orientation of the sequence is from SP6 to 7' end of the BAC
 clone.

Genes were identified by a combination of several methods: Gene
 prediction programs including GenScan and Genscan+ (Chris Burge,
<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/genemark/>), and Gensplicer
 (Mhaela Pertea and Steven Salzberg, [contact mpeete@tigr.org](http://contact.mpeete@tigr.org)),
 searches of the complete sequence against a peptide database and
 the plant EST database at TIGR (<http://www.tigr.org/tdb/tgl.shtml>).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as unknown proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as hypothetical proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are
 identified by RepeatMasker (Arion Smith,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES
 source
 1. 158714
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="3"
 /map="R1420"
 /clone="OSJNBa0013M12"
 complement(830..2323)
 /gene="OSJNBa0013M12.30"
 /note="predicted by genscan"
 complement(join<830..847,1073..>2323))
 /gene="OSJNBa0013M12.30"
 complement(join1830..847,1073..2323))
 /gene="OSJNBa0013M12.30"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAC6119.1"
 /db_xref="GI:12039332"
 /translation="MYDDDERRRGRRAPEALPDDLYVEITLRLAPSGRRARAK
 AMLSVTSQDPFLRAHARSTVAATVETRTARGSTTVCLVRQGRCSAV
 ASLAFSVASSEPSRMTVVIGWMDGLCAHLFGRGREVERVYLCPNLTETCTVP
 APSPTDGLVGVAHPPTTSRFHIMANFTTMEFTFWLIRLGNSVMREVRDALATMRV
 CTFELHAPVRLHGLCHMLTASSASSAOFLVAIVNMEEREERLEAPRGCGRGSSSH
 TMMGHTTHGHGICALADERPANALGMVLDDYSDPTSMRLRKIDYYSCGAGGAG
 AALDDPHAAAOTFRARFETADVVEPLRGVDDDBEGEITLQIGDEEVYVWGRRA
 WRKMGILLPTTRRLMHKOCILPREVLETV"
 complement(940..972)
 /rpt_family="AT_rich"
 complement(2590..2625)
 /rpt_family="AT_rich"
 2638..2720
 /rpl_family="Wanderer_O11 MITE element from gb:U34601
 Oryza sativa wanderer mobile element linked to Xa21 (411
 to 637) 227 nt"
 3300..4069
 /gene="OSJNBa0013M12.12"
 /note="predicted by genemark hmm"
 join(<3300..3458,3953..>4069)
 /gene="OSJNBa0013M12.12"
 join(3300..3458,3953..4069)
 /gene="OSJNBa0013M12.12"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAC6121.1"
 /db_xref="GI:12039334"

```
repeat_region /translation="MSKATRCGRDEADAVKPNKNDGSGRPQDERHGRGPRRRKPA
APPSPRRALVHSLEFDMWRFPGKEKLDSEFTFDRCCHKKGKSPDE"
complement(4241..4277)
/rpt_family="At_rich"
repeat_region /rpt_family="At_rich"
complement(4376..4417)
/rpt_family="(A)n"
repeat_region /rpt_family="(A)n"
complement(4765..4790)
/rpt_family="At_rich"
repeat_region /rpt_family="At_rich"
complement(4834..4871)
/rpt_family="At_rich"
repeat_region /rpt_family="At_rich"
complement(5839..5942)
/rpt_family="Crackie MITE element from gb:U72728 Oryza
longistaminata receptor-like kinase protein (Xa21), family
member F, pseudogene sequence (6201 to 6583) 383 nt"
5842..6255
repeat_region /rpt_family="Crackie MITE element from gb:U72728 Oryza
longistaminata receptor-like kinase protein (Xa21), family
member F, pseudogene sequence (6201 to 6583) 383 nt"
complement(6164..6239)
/rpt_family="Crackie MITE element from gb:U72728 Oryza
longistaminata receptor-like kinase protein (Xa21), family
member F, pseudogene sequence (6201 to 6583) 383 nt"
6438..7777
repeat_region /gene="OSJNBa0013M12.13"
/note="Predicted by genemark hmms"
join(<6438..7064,7130..>7777)
/gene="OSJNBa0013M12.13"
join(6438..7064,7130..7777)
/gene="OSJNBa0013M12.13"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG46124.1"
/db_xref="GI:12039337"
/translation="MEYDERRRRGRARRGALPDPLVEILLRLPAPSIGRCHAVCK
AULSTSGDPFLRAHARSCSTAVTAATVETRTTTCGRCSTTVRIIRLRCSGTA
VAASLAVSFVSASEPRSRKTAIVIGLMDGLCAHILFGRCGRYERVLNPLREKCTI
VPAPATDGLVGAGIAHPITSRFHIMHTNFETMETWILRLGENSVYRRTARISLIG
LVITGAVSNGRLQOEEMEREERLMEAPGQGRDNLSSHSTMGIHTHCGKLCA
LADIPGANAIGLVLEEDYSDPTSMRLQRIIDVYSCGSAAGALDDDDPHAAQGTFA
RSTADVFNGLNGVDDDEGEIILDLGEDEVYVNGRAAMRRGILPLTRCLMMH
RQCILPREVSPDSOVPEKDIGCHCFRIY"
complement(7812..7844)
/rpt_family="At_rich"
/rpt_family="At_rich"
complement(10397..11746)
/gene="OSJNBa0013M12.14"
/note="Predicted by genemark hmms"
complement(<10397..>11746)
/gene="OSJNBa0013M12.14"
complement(10397..11746)
/gene="OSJNBa0013M12.14"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG46127.1"
/db_xref="GI:12039340"
/translation="MDNAAPSSGAGCEMATAGALPDLDLDELILMPANSILCKR
AVCKAMRSRTHSHYFLRAHAARSRIIAAVVDTAIPDEFTTISIRLPGDGDGRS
GAAYSSSSSFASVRFSGRPVFLGSDVGCALPVAGVRLSRIDRVLNPILKACTS
VPPPATGIVIGYAHPTSRYLHLADADAPVNHMAATIRILRVGEKNWNRKJAR
HPAPACVESRTYIRREGARPVSLHGLILVAPLSAASAPRLISVDMRREERQDA
PROMARHGLPHMNGVQIARLSGLCAFPNERSASALGMWLTEDISDPSSMNERSID
YSRHGAGSRNIAKTERNRSAAATTAVEVLPDVGSGDVGGGGGEIIMPLFNQFTRE
AVYNGRAAMRRRIPLPPTRRYMAHKECMLEPREVSFGGAHVEESDIGHRCICI"
11941..15099
/gene="OSJNBa0013M12.24"
/note="predicted by genscan"
join(<11941..12207,13285..13390,14650..14753,
14896..>15099)
/gene="OSJNBa0013M12.24"
join(11941..12207,13285..13390,14650..14753,14896..15099)
/gene="OSJNBa0013M12.24"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG46129.1"
```

```
/db_xref="GI:12039342"
/translation="MGRSSRPAAOVTRSSPRRLPLRSISTSLAIRRRRRYPCRR
SPLAVAGMEGEENEYGPRTSRATRMATWASLASDNRRAOALRLALSDN
AGPEVVDLNDDEYGSADDEHDVIMORRKSMMKRTRGKLEKKAASFPDVLQEA
NIESLPPIVPTYSRAAVPPSTSSRRHNTYCGSTANTYCVACGRFCRCQVILND
TRCLKRVA"
complement(12036..12072)
/rpt_family="(CGG)n"
repeat_region /rpt_family="(CGG)n"
complement(15879..15898)
/rpt_family="(CGG)n"
repeat_region /rpt_family="(CGG)n"
complement(18634..18661)
/rpt_family="At_rich"
/rpt_family="At_rich"
19213..22689
/gene="OSJNBa0013M12.1"
/note="similar to pyruvate carboxyl peptidase
GI:3551425 (Pyrococcus furiosus): EST C19528, D23739,
AU094441, AU092288 from this gene"
join(19213..19412,20489..20806,21114..21218,21317..21364,
US-10-024-955-7 (1-213) x AC082644 (1-15874))

Alignment Scores:
Pred. No.: 5.44e+03 Length: 158714
Score: 95.00 Matches: 50
Percent Similarity: 42.33% Conservative: 30
Best Local Similarity: 26.46% Mismatches: 67
Query Match: 8.90% Indels: 42
DB: Gaps: 10

OY 42 GHSerGIuThrIleAspProMetLysValProAspHis-----AlaAspLys----- 57
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 147680 CAANAACAGCAATAGAAATATATCATCATCCCAACATTAATTCACAAATAATATGA 147739

OY 58 -----PheGIuATGHisValGIYIleValAspPheLysGIuLueu-AL 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 147740 TTAGTATCGCTTACCTTGATATATCATGCAGATGATGACAAACAAATGCTAGCTTGTC 147799

OY 72 aMetAgsAsnIleGIuAlaArgLysLysGIuMetLysArgGIuLncLysAspAlaAsnVa 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 147800 TCAAAACATGCTG-----AAGGAATTGGAAGATCAACATCCAGCAACAA 147844

OY 92 ILYsGIuGIuLncLysLysIleValLysAlaHisLysLysLysIleGIYValHisAspSpleVa 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 147845 GAATGGAAGAACTCAAGTGAATTCGACAAACAAATTCCAAGACTTCAACAGATATCAT 147904

OY 112 Iser-----MetGIuTyrAspLeuAlaTyrLysLysGIYAspLeuHisIspThrHisVa 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 147905 ATCTCATAGCGCTTCGGAAGCAGCTTAAGTTAGGA-----ATAGAACGTTCCATGC 147958

OY 131 ILeSerAspIleGIuAspPheValVal-----AlaLeuSerLysGIuLle-SerAspG 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 147959 ACAGAAAGACCTCCAGGAATTCCTGTAAATATCTCTCAACGTCAGATACAGAGATT 148018

OY 149 IuGIuAsnIleIleMetLysSerPheGIu-----ValArgGIuPheAlaAsnV 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 148019 CAGTAAATATATATTAATATGCTCCCTTCACAACTGTAGAAATGCAAA-----TTAAAT 148072

OY 165 aValAlaAsnHisIleGIYLeuSerLleLeuAspProIlePheGIYValLeuSerAspV 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 148073 CCACGAAAC-----ATGCACCATATTTTGCTCTTATTCATATA 148111

OY 185 aLeuThralaIlePheGIuAspThrValArgLysGIuMetLysValLeuAlaProA 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 148112 TATTTACTCTCTTTTTV-----TTTTAGTTCAGCTACCTT 148147

OY 205 IapheLysArgGIuLueuGIuLys 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 148148 CCTACAAAGGAATTCGAGGAA 148170

RESULT 22
ASNCALIA/C ASNCALIA 1266 bp mRNA linear PLN 06-DEC-2001
LOCUS Aspergillus aculeatus (clone pc1g1) endo-1,4-beta-galactanase
DEFINITION
```

```

(gall), complete cds.
ACCESSION      L34599
VERSION        L34599.1
KEYWORDS       arabinogalactan endo-1,4-beta-galactosidase; arabinogalactanase;
SOURCE         endo-1,4-beta-galactanase; galactanase.
ORGANISM       Aspergillus aculeatus.
REFERENCE      Aspergillus aculeatus.
AUTHORS        Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
TITLE          Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
JOURNAL        Christgau, S., Sandel, T., Kofod, L. V. and Dalboge, H.
MEDLINE        Expression cloning, purification and characterization of a
PUBMED         beta-1,4-galactanase from Aspergillus aculeatus
CURR. GENET.   27 (2), 135-141 (1995)
JOURNAL        95308535
PUBMED         7788716
FEATURES       location/Qualifiers
SOURCE         1..1266
                /organism="Aspergillus aculeatus"
                /strain="KSM510"
                /db_xref="taxon:5053"
                /cell_type="mycelium"
                /tissue_lib="YES 2.0"
                1..1266
                /gene="gall"
                28..1080
                /gene="gall"
                /fc_number="3.2.1.89"
                /standard_name="galactanase"
                /codon_start=1
                /product="arabinogalactan endo-1,4-beta-galactosidase"
                /protein_id="AA32692.1"
                /db_xref="GI:509837"
                /translation="MPASLLLAAPLTHAALTYRGADISSLLLEDEGYSYKLNQ
                TQALETLADAGINSINQRYWNPDSYDIDNLEAKRYKAGMSLYLDLHSDTW
                ADPSDQTPSGWSTTDGLTKWLYNTLVCNTFANDIDIELISIGNERAGLMP
                LGETSYSNIGALHSGANGVKDSNLAATTKIMHLDGDSMDQONTFETVATGEL
                LSTDEDFEGSVYPEYASATLASLKTSLANLOSTYKPVVVTNMPVSCPNPVAE
                PSDLSIPFVSAGQEFLEKLAIVENTTGLGVYVWPWIGNAGLGSSCADLAMD
                YTWDEYESTETGEL"
                28..75
                /gene="gall"
                76..1077
                /gene="gall"
                /product="arabinogalactan endo-1,4-beta-galactosidase"
                /note="bp 76-162: NH2-terminal residues"
BASE COUNT     285 a 342 c 352 g 287 t
ORIGIN
sig_peptide
mat_peptide
Alignment Scores:
Pred. No.:      22.6      Length:      1266
Score:          93.50     Matches:      52
Percent Similarity: 37.45%  Conservaive: 39
Best Local Similarity: 21.40% Mismatches: 77
Query Match:    8.75%     Indels:      75
DB:             8         Gaps:        10
US-10-024-955-7 (1-213) x ASNGALIA (1-1266)
QY 15 ValSerAlaAspProIleHisTyrAspIleThrGluGluIleAsnLysAlaIleAsp 34
Db 945 ATCAAGTACTGCGCTCCACACGCGTGCACACTTCGAGGAAGTCTCTGCGCCGCGAC 886
QY 35 ---AspAlaIleAlaIleGluInserLuhThrIleAspPrometLysValProasp 53
Db 885 CGAAGAGGATCGATCGAGATCCGAGGAGATCCGCGGCTTCGGGACAGAC 826
QY 54 HisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyIleuAlaMet 73
Db 825 CGGCGAGTTCCTTCACACACACCGGCTTGC-----GTAGTTCGACTGAGGTT 772
QY 74 ArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGln-----Gly 88

```

```

Db 771 CGCCAGACTAGCTCTTCAGGATGCCAAGTTGCCGACGCGCTGAGAACGGGTAGTACGA 712
QY 89 AspAlaAsnVal-----LysGlyGluGluGlyIleVal 99
Db 711 GACACCAAGATGATGTCGAAGTCCGTCGACAGCAGCTCCGCCCTGGCCAGAACGGTCTGTA 652
QY 100 LysAlaHisLeuIleGlyValHisAspAlaIleValSerMetGluTyrAspLeu--- 118
Db 651 AAAGTAGTTCGTGATGATCCAGCT---CCACCGCGTGCAGAAATGATGATGATCTGGG 595
QY 119 ---AlaTyrLysLeuGlyAspLeuHis---ProThrThrHisVal----- 131
Db 594 GGTGTCCGACGGTGGTAATCTTCACATCCCAAGCACCAGCTGACAGCTGCCGAT 535
QY 132 -----IleSer 133
Db 534 GTTCGAATAGCTGCTGCTTTCGCCACAGGCCAGACGACCGGACGATCTCATTTGCC 475
QY 134 AspIleGlnAspHe----- 138
Db 474 GATCGAGATGATTCGATATCATATGCTGTCGCAAAAGTGTGATCTGATGCTGACGCT 415
QY 139 ---ValValAlaLeuSerLeuGluIleSerAspGluGlyAsn----- 151
Db 414 GTAGTTGTACAGCTGCCATTTCAAAAGTCCGAGATCGCTGTCGACACGCCGGAAGTGT 355
QY 152 -----1LeThrMetThrSerPheGluValArgGlnPhe 162
Db 354 CGTGTGCTGCTAGAGATCCGCCCATGTGTCTACATGAGTGCAGGTGCAGTAAGATCAT 295
QY 163 AlaAsnValAlaHisHisIleGlyIleuSerIleLeuAspProIlePheGlyValLeu 182
Db 294 GCCAGCCGCTTGAACCCGCTTGGCCAGCTCCAAATTTGATGATGATGATGATGATGAT 235
QY 183 SerAspVal 185
Db 224 GCTGGGCTT 226
RESULT 23
CNSPAX02/c 293250 bp DNA linear BCT 09-SEP-1999
LOCUS        Pyrococcus abyssi complete genome; segment 2/6.
DEFINITION   Pyrococcus abyssi complete genome; segment 2/6.
ACCESSION    AJ248284 AL096836
VERSION      AJ248284.1 GI:5457730
KEYWORDS     complete genome.
SOURCE       Pyrococcus abyssi.
ORGANISM     Pyrococcus abyssi
              Archaea; Euryarchaeota; Thermococci; Thermococcales;
              Thermococcaceae; Pyrococcus.
              1 (bases 1 to 293250)
REFERENCE    Hellig, R.
              Pyrococcus abyssi genome sequence: insights into archaeal
              chromosome structure and evolution
              Unpublished
              2 (bases 1 to 293250)
GENOSCOPE    Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
FEATURES     location/Qualifiers
SOURCE       1..293250
              /organism="Pyrococcus abyssi"
              /strain="Orsay"
              /db_xref="taxon:29292"
              /complement(115..1857)
              /note="PAB148"
              /codon_start=1
              /transl_table=11
              /product="ABC TRANSPORTER ATP-BINDING PROTEIN"
              /protein_id="CAB49221.1"
              /db_xref="GI:5457731"
              /translation="MADKSLSLKRFIKELSERRTLAIVVSIYGSALANLASPYI

```

LRVAIDHYIIPGRKELGFIAMLYILSLVAQMFETTIDQFTYEVFGQVLRNLRSQLH
EKLIQSLNDFERTKSTGDLVSRIINDTNVNDVLSGLGSLSLAGIIVAMLL
DVKLTIVLTSPVPLAVIIALYFGKMRARVETRQIKAKISSVVESSVAGIETIKAFG
KEKVEFEKSKVETIKAYLRVAIYMGIFMPMIASILSYIVIVAYGVALYAGV
SIGVVAFTIOVAORRCPINNVSMYDLSGLALBERIYELIDOEENEDEGISVEK
LKGEIEFRNWFVEYKGRPVLDVNLVIPPASKVAIVKGTGAKTTIANLIRFDPPT
KGEIYIDGIDGKISRKLKRIGVPOETIYLPETIMENILMANPEKKEDEIVKCK
ELGIHDFIMKLPKGYETISAGAGKLLSVGERQLSLARAMEKDPDIVILDELSVDP
KTEQLVQNAAMIKMEGRTSIIIAHRLSITRERADKVIYVDDGRIVEGSGPSELLSRKY
FYKLYTSQMOKEVPG"
complement(1847. .3586)
/note="PAB2147"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="CAB49222.1"
/db_xref="GI:5457732"
/translacion="MRKSIINQILRLSLRGHAHFILGLFIILMSTNSIVPLIR
NAIDGVSTGAKVARIYGLLLLVGLVNGVFSLRGLLVKAAQVAVYILMDRFA
IQRHMEFPDFTFESQLSRITNDPERITRFLSPVRLYSIFLIISLYVMRMI
TLVMVALITIAIVGLNTTYAMKVRPIYDKIRHOTGLIASIATGTIAGITIKISVE
ENVOTKFSOENELYSLVNKKAKITAYGNAPFVLGLAMSTMEFYGGKAIIONTLV
GELVAFIYMLTMMPLRALGFTIGDIOBSLAASRLPDDIVDSAPKSIDPPAVESIN
PGEIEFRDVMITITGRTVYLGVSFKVPGCKVYITTPSGSKSIIKLIRAPFT
KGEVLIDGVNAKITEHDRKIYAVIPEPTEFNRSIRENTALAKPNATQEBVERAK
IAKIHDFIASLPKGYDVGEGITLSGGORIALARAMEKDPDIVILDDPSMDA
ETEKVLDEKLKELQDKTALIVSQRLSLVKLVDRVIVMKDKRIVEDGKPEELIKRGI
FSEMKACGVYGG"
3753. .4451
/gene="succ-1like"
3753. .4451
/gene="succ-1like"
/note="PAB0203"
/codon_start=1
/transl_table=11
/product="SUCCINYL-COA SYNTHETASE, BETA SUBUNIT related
protein"
/protein_id="CAB49223.1"
/db_xref="GI:5457733"
/translacion="MDRIAKAREIIEKAKAENRPLVEPEAKEILKYGPIPEKVAR
DEEAVKFSREIGYVPMKIVISPOIHKSDAGVKINIKNDEAREAFRTIMENARY
KPDADMGVIYRMPLGKEVYGMIRPOGRPAVPGIGGFVEILKDVSRVAPIS
KEDALEMKETIKAYPIILAGARGKFPVNIDALDIYKGEALDELPEIRELDINIFA
YDEAVAVADARLL"
4543. .4878
/note="PAB0204"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49224.1"
/db_xref="GI:5457734"
/translacion="MEDENSFMRLFRKKNGCILBOEINEDVYKELSLMRAP
EYIGEKIVLHPQADVILIPRLIYKASRREAKELINLNLHHYSQPPGLMPAYGLYS
IRKEKGRVS"
4960. .5688
/note="PAB0205"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49225.1"
/db_xref="GI:5457735"
/translacion="WNISEAIVFGLGLVLSLVRIRKFGSGIRFALLIIVLLGID
VFLKGEMLVNHSIVSLFGLVFLISGLKHLDEPKRVQOQVYVNMGENSGF
VGAITIPGSRSDIYDLNIMIKELNAPVLFYTFYPSFYNLGENIRVWITQSEGGIP
PKTHIYDGYAIRFARENKNAVITIIDCLEYILLYNEFSSVFETLANLMDHLVNMNSAL
ILAVQKALETYIYMLNNEPPL"
5683. .7016
/note="PAB0206"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49226.1"
/db_xref="GI:5457736"

/translacion="WISPSKSHSLENKRYKLEKRIIYLLIKRKLIGIPLVILFGLV
LAFYIASEEDNEVEFFYNSFAWERTITLEDKGNAPIEITGDSNGIDPLIVSGDV
SLVIGRGEYVIFGIPVGSSEIVATYLUKSKVOISGLRLOQNTYLVVERGSOKVILNPG
EISYISVSKSRSSNSVILYVDEAGRHNITLIVRENNKMSSTYRLILKDEGVEGL
AIINENPSKARLKRGYIYVADIGGTYPVYLAFTMOAGAGEYIKHEHTLLEPQNEV
NGESVNAIKETESLEMERSTLYESMPQOEGPIYEVIRFTDKPLPQNGEVYREED
GKLLIRLISLQOKASGVYVEIKIGRETELGGTKVIVORVNEGDKTKVETITLENLS
NEPREVMIKHKHIGRITYSLNPLEETADVILEKVNPDGKEVNNLEYET"
complement(6994. .7917)
/note="PAB2145"
/codon_start=1
/transl_table=11
/product="UDP-glucose 4-epimerase (gale-1)"
/protein_id="CAB49227.1"
/db_xref="GI:5457737"
/translacion="MRKLIYVAGAGFISGSHIAEALKEENDVYVINDLYSGKPEVNP
EGVRIEADVDRDYESIARIETVSSADVVPHEAIOISVEESVQDPVPEEVNVTGIIINIR
ALSENGKLIIFASSNAVYGEPTSLPIREDHNLNPSIPGYSKYSEHCKYFQLOXG
PTVILIRYVNGPROSSAVAGVISIFEMRALRGEPLVIFGQKQSDPEYVYKDVVEAN
ILVAKERRAEGETFNVGRETITIELMKIIELSSTSQILFSKPRGDIRSVADI
EKIKRIGFRPYSLEEGHLETWKMKSRITG"
7973. .8632
/gene="gph"
7973. .8632
/gene="gph"
/note="PAB0207"
/codon_start=1
/transl_table=11
/product="PHOSPHOGLYCOLATE PHOSPHATASE"
/protein_id="CAB49228.1"
/db_xref="GI:5457738"
/translacion="MKIKGILFVDDETLLEYEGYSLRMYENVGLIPANKKILGLVDWE
TFRKMAKGEIPRSYVEKENPFDHEFKALDRANRERRLRBEKVHVFDDVYALKE
KLGIKILAAVNSAQDNTELVLKAFDLKTYDVYGVKQSYLDGKPPMPYILINKALRA
LNIRKEAAILVGDSELDYRAGKNAVAVIQLVREKVVAGADMYIKTLMELVDLLKQV
L"
8973. .9578
/note="PAB0208"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49229.1"
/db_xref="GI:5457739"
/translacion="WGEELMDLVDVNETRRRIILLITKRPYVSELSQELGYGQKA
VLEHLRIIEEAGLIESRIEIKIPGRPRRYVMIKRLRIEILITPTLFGSEMYEPRDIR
KSPEYQARLELIKSTPEIDVKKMRELAELHDEIKIEIKELRELEVKILYERYIE
NVMRMSSENRQTEKIKFEIERLIPROYARIRKKEFTENT"
9610. .9894
/note="PAB3084"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49230.1"
/db_xref="GI:5457740"
/translacion="MKRLGKVSYHAKQGLLVYRSTWPSLNDPVIDKRLFYGIKDV
FGPVKRPYVAIKPVDDPEKYGVQLYIDERKKKKRKKRGKMKKFKFG"
9920. .10411
/note="PAB0209"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49231.1"
/db_xref="GI:5457741"
/translacion="MLIGVLSDTHPKRAYFPDRLVLPFEKKVKYIIHAGDITEKQOL
DLLESVAVIYVKNAGARLIDPEERTLKVQKILVILGHNPLSLDQNLVYKALEED
ADILIFGHTHRPYNKITAMGKEYVLLNPGSPTLPRMSEPTVAILNVRGIDIVTFHY
WKT"
complement(10408. .11010)
/note="PAB2143"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49232.1"

```

/db_xref="GI:5457742"
/translation="MIKFVASEALICPRRWFRIRKPEEMPEVAKPRLKGVIEHET
LGRFLKEFNVLEKSVILRLPKLMEKTHGIDVPIEIKSVKPYPIEYHILA
OLNLYRGEVADYGLYIKIGDPMKIIINLEDEPFIYGQNVKVEYVDVLELRF
ETIKYFIEKIFLEEDKIPAMRGPHCNFYKILCONTL"
CDS
11334..11912
/locus_tag="PAB0211"
/codon_start=1
/translation_table=1
/product="hypothetical protein"
/protein_id="CAB49233.1"
/db_xref="GI:5457743"
/translation="MWIIPRIDPRIDIKIRKELGITOEELARRAGVQAYIAKLEAG
KVDRLSTFNRIALLLECOKRTITAKSINSPISAPEDIIDIVYMEKYNISOV
PVIYKDVGVAITERLLVRSKLEDEDYSKRAVIMESPFVSEDEDELYIKYLLBE
HPAYIVGCKNKPIGITIRSDIFKGRTEKE"
11964..13409
/gene="soxa"
/locus_tag="soxa"
11964..13409
/gene="soxa"
/feature="PAB0212"

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 1.53e+04	293250	60	35	74	87	14
Percent Similarity: 37.11%						
Best Local Similarity: 23.44%						
Query Match: 8.75%						

US-10-024-955-7 (1-213) x CNSPAX02 (1-293250)

```

Oy 2 MetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIleHis 21
Db 80291 CTCAAAGCGATGACCGGTGACGGGGGACGGTTCTGTCGAGCTG---GCAAAAGCCGGAAGAG 80235
Oy 22 TyrAspLysIleThrGluGluIleAsnLysAlaIleAspAlaIleAlaIleGlu 41
Db 80234 TTCGAAAGCTGCGTGCAGGAATTCAG----- 80208
Oy 42 GlnSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGluArg--- 60
Db 80207 -----CTTGAAACGTATGCCACGAGAGTACTGCTGCGCAGAGA 80169
Oy 61 -----HisValGlyIle----- 64
Db 80168 ATTATGGCAGCTCAAGCATCTCTGGCGGCGCAACAGATGGAGTACTCCCGAGGAAG 80109
Oy 65 -----ValAspPhe-----LysGlyGlu---LeuAlaMet 73
Db 80108 CACTTACCGGAGCGGAGTACATGCGTACGCTGTTAGCGGAGAGAGCGTTGCCCTC 80049
Oy 74 Arg---AsnIleGluAlaArgGly-----LeuLys 82
Db 80048 GAGTATACGTCGTAAGACGAGCGGCTTTCTTCTGTGCTTCCCTGACAGAGCGGTGTC 79989
Oy 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGlu-----GluGlyIle 98
Db 79988 CTCATTAAGAGGCGGAGTACGACCATATGAGAGAGGCGCCCAATATTAAGCCCTTA 79929
Oy 99 ValLysAlaHisLeuLeuIleGlyValHisAsp---AspIleValSerMetGluTyrAsp 117
Db 79928 GCTGGATGAGCGGTAGATTTGGGAATAGTACCTCGACCGCGCCCTTATCCCTTAATA 79869
Oy 118 LeuAlaIleTyrLysLeuGly---AspLeuHisProThrHisValIleSerAspIleGln 136
Db 79868 CTAATTAATGACCTCGGACCTGATTAACACAGAGTGAAGTATTCATGAGCTA--- 79812
Oy 137 AspPheValAlaIleLeuSerLeuGluIleSerAspGluGlyAsnIlePheThrSer 156
Db 79811 -----ATGGAAGCTTACGAAAGGAGGATTAATTAAGAGC----- 79779
Oy 157 PheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAsp 176

```

```

Db 79778 -----GACGAGATAGCGGCGCTAAAG----- 79758
Oy 177 ProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLys 196
Db 79757 CCGGTGGGAGATATGAAACGGTCTCTCACTCAATGAGAATATAGCTACGAGAAG 79698
Oy 197 GlnMetThrLysValLeuAlaProAlaPheLysArgGluLeuGluLys 212
Db 79697 GCGCTTGGTGAATGTTCTGCTACGCGGCTCGAAGCAGCAGTGAAGAG 79650
RESULT 24
AX041918/c 349980 bp DNA linear PAT 23-NOV-2000
LOCUS Sequence 813 from Patent WO0065062.
DEFINITION AX041918
ACCESSION AX041918
VERSION AX041918.1 GI:11340677
KEYWORDS
SOURCE
ORGANISM
Pyrococcus abyssi.
Pyrococcus abyssi
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE
1 (bases 1 to 349980)
Forterre, P., Thierry, J.C., Prieur, D., Dietrich, J., Lecompte, O.,
Querellou, J., Weissenbach, J., Saurin, W., Hellig, R., Flament, D.,
Raffin, J.P., Henneke, G., Gueguen, Y. and Rolland, J.L.
Genome sequence and polypeptides of pyrococcus _labyssi_, fragment
and uses thereof
Patent: WO 0065062-A 813 02-NOV-2000;
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFRIMER
INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)
FEATURES
location/Qualifiers
1..349980
/organism="Pyrococcus abyssi"
/db_xref="taxon:29292"
/feature="Original length of seq 1: 1765118 (for 812
sequences) New sequences : seq 001, from 0.000.001 to
0.349.980, length: 349.980 seg 813, from 0.300.001 to
0.649.980, length: 349.980 seg 814, from 0.600.001 to
0.949.980, length: 349.980 seg 815, from 0.900.001 to
1.249.980, length: 349.980 seg 816, from 1.200.001 to
1.549.980, length: 349.980 seg 817, from 1.500.001 to
1.765.118, length: 265.118"
BASE COUNT 92953 a 77841 c 81831 g 97355 t
ORIGIN
Alignment Scores:
Pred. No.: 1.89e+04 Length: 349980
Score: 93.50 Matches: 60
Percent Similarity: 37.11% Conservative: 35
Best Local Similarity: 23.44% Mismatches: 74
Query Match: 8.75% Indels: 87
Gaps: 14
US-10-024-955-7 (1-213) x AX041918 (1-349980)
Oy 2 MetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIleHis 21
Db 87391 CTCAAAGCGATGACCGGTGACGGGGGACGGTTCTGTCGAGCTG---GCAAAAGCCGGAAGAG 87335
Oy 22 TyrAspLysIleThrGluGluIleAsnLysAlaIleAspAlaIleAlaIleGlu 41
Db 87334 TTCGAAAGCTGCGTGCAGGAATTCAG----- 87308
Oy 42 GlnSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGluArg--- 60
Db 87307 -----CTTGAAACGTATGCCACGAGAGTACTGCTGCGCAGAGA 87269
Oy 61 -----HisValGlyIle----- 64
Db 87268 ATTATGGCAGCTCAAGCATCTCTGGCGGCGCAACAGATGGAGTACTCCCGAGGAAG 87209
Oy 65 -----ValAspPhe-----LysGlyGlu---LeuAlaMet 73

```

Db 87208 CACTTCACCGAGCCGGAAGTACGCTACCTGCTTACGAGAGAGGCTTGCCCTC 87149
Qy 74 Arg---AsnllleGlualaArgly-----LeuLys 82
Db 87148 GAGTATACGTGACGACGAGGGGCTTTCTTTGTGTGCTTCCTGCACAGCGGTTC 87089
Qy 83 GluMetLysArgGlnGlnLysAlaAsnValLysGlyLg-----GluGlyIle 98
Db 87088 CCGATGAAGAGGGGGAAGTACGACGAGGAGGCCCCGCAATATGAAGCCCTA 87029
Qy 99 ValLysAlaHisLeuLeuIleGlyValHisAsp---AspIleValSerMetGluTyrAsp 117
Db 87028 GCTGGGATGACGGCTTAGATGGGAATAGTACCTCGACCGCCCTTATGCCGTAAA 86969
Qy 118 LeuAlaLysLysLeuGly---AspLeuHisProThrThHisValIleSerAspIleGln 136
Db 86968 CTAATTAATGACCTCGACCTCGACGTGATTAAGCAGAGTGAAGTATTCATGCTA--- 86912
Qy 137 AspPheValAlaAlaLeuSerLeuGluIleSerAspGlnGlyAsnIleThrMetThrSer 156
Db 86911 -----ATGAGAGCTCTACGAAAGGAGCATATTAAGAGC----- 86879
Qy 157 PheGluValArgGlnPheAlaAsnValAlaAsnHisIleGlyLysLeuSerIleLeuAsp 176
Db 86878 -----GACGAGATAGCGCGCTTAAAG----- 86858
Qy 177 ProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLys 196
Db 86857 CCGGTGTGGGAGATATGGAACGGTTCTCAGCTCATAGACATATAGCGTACGAAAGG 86798
Qy 197 GluMetThrLysValLeuAlaProAlaPheLysArgGlnLeuGluLys 212
Db 86797 GCGGTGTGGATGTCTCGCTGACGGGCTCTGACGAGCATGAAG 86750

RESULT 25
AE010403 10558 bp DNA linear BCT 01-AUG-2002
LOCUS Methanopyrus kandleri AV19 section 102 of 157 of the complete
DEFINITION genome.
ACCESSION AE010403 AE009439
VERSION AE010403.1 GI:19887697
KEYWORDS
SOURCE Methanopyrus kandleri AV19.
ORGANISM Archaea: Euryarchaeota: Methanopyrii: Methanopyrales:
Methanopyraceae: Methanopyrus.
REFERENCE
AUTHORS 1 (bases 1 to 10558)
Slesarev,A.I., Mezhevaeva,K.V., Makarova,K.S., Polushin,N.N.,
Shcherbina,O.V., Shaknova,V.V., Belova,G.I., Aravind,L.,
Natalie,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O.,
Malykh,A.G., Koonin,E.V. and Kozayvkin,S.A.
TITLE The Complete Genome of the Hyperthermophile Methanopyrus kandleri
AV19 and Monophyly of Archaeal Methanogens
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10558)
Slesarev,A.I., Mezhevaeva,K.V., Makarova,K.S., Polushin,N.N.,
Shcherbina,O.V., Shaknova,V.V., Belova,G.I., Aravind,L.,
Natalie,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O.,
Malykh,A.G., Koonin,E.V. and Kozayvkin,S.A.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2002) Fidelity Systems, Inc., Gaithersburg, MD
20879

FEATURES
source 1..10558
Location/Qualifiers
/organism="Methanopyrus kandleri AV19"
/strain="AV19"
/db_xref="taxon:190192"
complement(196..1083)
/gene="MK1131"
complement(196..1083)
CDS
/gene="MK1131"
/codon_start=1

/transl_table=11
/product="Uncharacterized protein specific for M.kandleri,
MK-23 family"
/protein_id="AA02344.1"
/db_xref="GI:19887698"
/translation="MSDAHELRRLRLKLVVDLSLKLRLKEGSSYANVRNPRLGLYLKLA
EEELVEPRVNAHPLVTPPISTVLRLELESESEVGYEVTASVEVAEDDGRVEEVEV
VDREVPVILGLPGVLTLSDRVEELASERAPLELDGMRLHPRAVPSVLDLTPPLAA
LRAVESVSGEDAFERVAVARIGHGLTVDRGVRSGCGGLKLVDTVDLRLASVLD
LARYVPAALPGVTVRELERLRLDGFDEREDLTBARAFELLSNLTPGQREEDVKIP
EHPSTIREVRLRDLYE"
2284..2691
/gene="MK1132"
2284..2691
/gene="MK1132"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein"
/protein_id="AA02345.1"
/db_xref="GI:19887699"
/translation="MDVTEELVRMAEKLSEARALEEGESPAAVAHAQVELLVK
AALRELVDPEPRTHDIRLGLVTLTLELDLGDGLAVEVRIAREHRTLASLLEGYE
SLFEPFEGDAEEVYRAKGVREDLRILEKYR"
complement(2904..3134)
/gene="MK1133"
complement(2904..3134)
/gene="MK1133"
complement(2904..3134)
/transl_table=11
/product="Uncharacterized protein"
/protein_id="AA02346.1"
/db_xref="GI:19887700"
/translation="MDTSGHRLVLRVYKVPFAPMAHARGRIIMGPGTIVRTSGRMI
LYFAYPIRHALGALALALQWNAVRAAMEE"
6088..6945
/gene="MK1134"
6088..6945
/gene="MK1134"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein"
/protein_id="AA02347.1"
/db_xref="GI:19887701"
/translation="MPEYPKPCTVREAVYVLSVPGDRATLIDLPDVPMPARVLDLYLS
AGSRVIGAMVRSAYSGLDHVPAPLELGFYRLDAGERTERMSDEPPTDPLRG
ILKRGLRWTVVVRFLPRLNGLEMAFCVPLDGFVMPAPFGVYVYGGVTEALVP
EGCPRLRITFGLPLPSLDLDTLVLYVLDGPRPRAVYVSSGFRHTYLVFRPVGASPEAYR
LADPLRVALLGSAVSLVPRVTKALALLVPAAGLDYRMPMLTGMATLGPAY
LFLCLLHLV"
6996..8108
/gene="MK1135"
6996..8108
/gene="MK1135"
/codon_start=1
/transl_table=11
/product="Permease of the major facilitator superfamily"
/protein_id="AA02348.1"
/db_xref="GI:19887702"
/translation="MPGRYPKDVLLGAVSLNDVSSGICFALLPYITITAGCGPALV
GGTWGEMELVKSLLNVASGRVGRGMAKPAVAAGYGSACKLLALLAVRDVYTGVLV
ASLERVKGKIRTPPRDAILAGLSEPGVLEVRNALDTSGAVLLAGLVLTIVGVV
AATAVLAELVGLFLSVPLIPEERLEEGSSGEGRGPRPALATVIGVLRGAVGMM
YSLRVLGVGGPAAAFVAGFSTLHALTSVPSGRISDRGPRGVALCIGYAMTAAALL
AGSLAVAPMLYGLAYGVDAVERAAVAELISCTAASYGAYHALVGSIVLGLVYVG
IMESVSPNAEYSACLGVAAALAPVILIRGSGC"
8288..8917
/gene="MK1136"
8288..8917
/gene="MK1136"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein specific for M.kandleri,
MK-1 family"


```

/Protein_id="AAM02349.1"
/db_xref="GI:19887703"
/translation="MPELVESHVHVPIPSDVERIRALRVAREVYRRKKPSLEYTRY
TVGSGTCPTTVARRKRDSEKRGHTLYLDKSENSVSFEWLVSDSEVLELARHL
MNRKSLVLTLEVLTLPLKRRARVLRGLALAFDAPSPRIIDLLELPORLES
FAVRILGMPAHVSSYLKRVIRSRRLSDGRHEVDVLELELRLRHG"
complement(8959..9327)
/gene="MK1137"
complement(8959..9327)
/gene="MK1137"
/codon_start=1
/transl_table=11
/product="Predicted nucleotidyltransferase of the DNA
polymerase beta superfamily"
/protein_id="AAM02350.1"
/db_xref="GI:19887704"
/translation="MSGDLVLEARINRRYREFVSRMREMERIARVARLLGEGT
RVFVGSVAKGEALPGSDVLIVSDGVPEDLELDREFERELDPERHPTEFHL
ATPEYEVNRLDEVEEF"
9348..9569
/gene="MK1138"
9348..9569
/gene="MK1138"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein"
/protein_id="AAM02351.1"
/db_xref="GI:19887705"
/translation="MRLRKRLERAEGSLEAARWLLRGNTDLALVMAERAGLRIR
AAQILAGEHARGHDLRDLRGARRHRRV"
9562..9735
/gene="MK1139"
9562..9735
/gene="MK1139"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein"
/protein_id="AAM02352.1"
/db_xref="GI:19887706"
/translation="MCEHAKALDVLGVATYTCRGAMEPSEETLEMAIGTVEELFELL
ERVERHGPASRP"
complement(9798..10037)
/gene="MK1140"
complement(9798..10037)
/gene="MK1140"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein"
/protein_id="AAM02353.1"
/db_xref="GI:19887707"
/translation="MIRAVVGTGVAATVGLSLPPYCGMLAPAMLIGFLTLKTERM
VRDRHIVGMGRVETVRYTPWPSRSTRSAPGGP"
complement(10034..10474)
/gene="MK1141"
complement(10034..10474)
/gene="MK1141"
/codon_start=1
/transl_table=11
/product="Uncharacterized membrane protein specific for
M.kandleri, MK-17 family"
/protein_id="AAM02354.1"
/db_xref="GI:19887708"
/translation="MVALIAPQLVVTGTAAGLFIHAMERKKRAMDVLRGKVEE
KKPRSWEPERFKNSRCDPVRILVFRPLINCAOSGYLLVSTMGWALATFYEMSVL
LGGHSPSPVRELVSFGLLVFSRSDVALLPQFRSPSRGS"
3304 C 3472 G 1935 T
BASE COUNT 1847 A
ORIGIN

```

Alignment Scores:
 Pred. No.: 349
 Score: 92.50
 Percent Similarity: 43.05%
 Best Local Similarity: 23.77%

Length: 10558
 Matches: 53
 Conservative: 43
 Mismatches: 68

```

Query Match: 8.66% Indels: 59
DB: 1 Gaps: 15
US-10-024-955-7 (1-213) x AE010403 (1-10558)
QY 16 SerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAsp 35
DB 2083 TCTCGGACCCG-----GAAACCTACCGCGGACGTCGCCCGACCTCAGGAG 2133
QY 36 AlaIleAlaIleGluGln-----SerGluThrIleAspPro 48
DB 2134 GTTTGAGAGCTCTCGAAGCGCGCTGACAGCGCGCCCGATAGAGTGAAGACCGGATCT 2193
QY 49 MetLysValProAspHisAlaAspLysThrGluAlaGlnHisValGlyIle----- 64
DB 2194 GTGGAGTCCCGGACACCTCGACAGGTGCAGTGCATGTGGGGGCGCTCCGGACGCG 2253
QY 65 -----ValAspThrLysGlyGluLeuAlaMetArg 74
DB 2254 CTCGCCGTGGGCTACCGGTTCCGGAGGATGACAGCTGCACAGAGAGCTGTCGG 2313
QY 75 AsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGly 94
DB 2314 ATGCCCGAGTGGAG-----CTCTCGAGAGCGCGAGCG-----GCCGTGAGGAGGCG 2361
QY 95 GluGlu-----GlyIleValLysAlaHisLeuIleGlyValHisAspAspIleVal 112
DB 2362 CCGGAGTCCCGCGCGGTCGCGGCGGATCGATCGAGCTCTGTCGTCGTCGTCGTCGTCG 2421
QY 113 SerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIle 132
DB 2422 CTTCGGAGCTGACGTGAGTGCAG-----CCACGAGAGCTGAC----- 2457
QY 133 SerAspIleGluAspPheValValAlaLeuSerLeuGluIleSerAsp-----GluGly 150
DB 2458 ---GATATCCGACGCTCTCGCGCTGCTACAGAGGAGCTGCTGCGGCGGAGG 2514
QY 151 AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHis----- 168
DB 2515 GAGCTCGCGGCTC-----GAGTCCGACCGAGTCGAGG-----GAGCACCGCGGAC 2559
QY 169 -----IleGlyLysLeuSerIleLeuAspProIlePhe---GlyVal 181
DB 2560 ACGTGGCGTCTCTCTCGAGGGGTACTTCTCCGTTCGACGCCCTTCGAGGAGAGC 2619
QY 182 LeuSerAspValLeuThrAla-----IlePheGlnAspThrValArgLysGluMet 198
DB 2620 GCCGAGAGGTCTGTCGCGGCGCGCGGAGGAGTCTTCGAGAGC---CTCGCGGATCTCTG 2676
QY 199 ThrLysVal 201
DB 2677 GAGAAGGTA 2685
RESULT 26
SC135/c 40909 bp DNA linear BCT 12-MAY-2002
LOCUS SC135 Streptomyces coelicolor cosmid 135.
DEFINITION AL031541 AL645882
ACCESSION AL031541.2 GI:20520886
VERSION 3
KEYWORDS 3-hydroxyacyl-coA dehydrogenase; 50S ribosomal protein L20; 50S
ribosomal protein L35; arabinosidase; argC; argJ; ATC initiation
codon; ATP/GMP binding protein; dehydrogenase; glutamate
N-acetyltransferase; glutamine synthetase; hncC; IS117; mult
domain; N-acetyl-gamma-glutamyl-phosphate reductase;
phenylalanyl-tRNA synthetase; pheS; pheT; rplI; rpmI; rRNA
methylase; secreted serine protease; transcriptional regulator;
translation initiation factor IF-3; transposase; TTA Leu codon;
two-component sensor.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 40909)

```

AUTHORS redenbach,M., Kieser,H.M., Denapate,D., Etcherner,A., Cullum,J., Kinasli,H., and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
PUBMED 8843446
REFERENCE 2 (bases 1 to 40909)
AUTHORS Oliver,K. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 40909)
AUTHORS Parkhill,J., Barrell,B.G. and Rønjandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT On May 9, 2002 this sequence version replaced gi:3581833.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/projects/S-coelicolor/) CDS are numbered using the following system eg SCDB7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., (Gene 30:157-66(1984)) as implemented at http://www.nih.gov.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid i35 lies between l10 and l41 on the AseI-I genomic restriction fragment.

FEATURES

source location/Qualifiers

1..40909

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid I35"

complement(548..1576)

/gene="SCO1580"

/note="argC"

complement(548..1576)

/gene="SCO1580"

/note="SCOI580"

reductase, len: 342 aa; identical to ARG_C-SRRCO N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (249 aa). Contains P501224 N-acetyl-gamma-glutamyl-phosphate reductase active site. Referred to as 'argA' on genetic map."

/codon_start=1

/trans_table=1

/product="N-acetyl-gamma-glutamyl-phosphate reductase"

/protein_id="CAA20791.1"

/db_xref="GI:3581835"

RBS		complement(559. .562) /gene="SC01580" /note="possible RBS upstream of argJ"
misc_feature		complement(830..1876) /note="99.48 match to SCARGC X66783 S.coelicolor argC gene (partial)"
misc_feature		complement(1106..1156) /gene="SC01580" /note="PS01224 N-acetyl-gamma-glutamyl-phosphate reductase active site"
RBS		complement(1584..1589) /note="possible RBS upstream of argC"
gene		1610..2173 /gene="SCI35.03" /note="SC01581"
CDS		1610..2173 /gene="SCI35.03" /note="SCI35.03, putative membrane protein, len: 187 aa. contains possible hydrophobic membrane spanning regions" /codon_start=1 /transl_table=1 /product="putative membrane protein" /protein_id="CAZ0792.1" /db_xref="GI:3581836" /db_xref="SPTREMBL:O8804.1" /translation="MPMHVYAMRCGYSRRCRRRLACRFPTSPGVAVRAMIVELV VGISRMQSAALIVSLCALPIFMFVGNTDPRGCGTAAVTEFGDCMDPCGAGML ADGSGRRIRIGDLFTAEELVEDIPLEFGGRDAVADRGTLIRNAVAQIYAAPGAAL VRIARRASINQETALMRRYRTTP"
gene		complement(2163. .2687) /gene="SCI35.04c" /note="SC01582"
CDS		complement(2163. .2687) /gene="SCI35.04c" /note="SCI35.04c, ATP/GTP binding protein, len: 174 aa; similar to TMRB_BACSU tunicamycin resistance protein (197 aa), fasta scores; opt: 282 z-score: 488.2 E(): 6.1e-20, 39.5% identity in 119 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop)" /codon_start=1 /transl_table=1 /product="ATP/GTP binding protein" /protein_id="CAA20793.1" /db_xref="GI:3581837" /db_xref="SPTREMBL:O8804.2" /translation="MTWVLNSTHGAKGTTCALIQRLLPSSRVFDARKVGETIMDTIP GILPDNDNHKMPRPVLEVTAARVLDYTGGTLVMPITVLEEYTRKEIAGGLAHGPV VRHFLVADDOHPLRIGRIAGDVLPDGDPSEFRLLPEYPAEAARTWLHAEEAELVDTHLTLP PAEALRIATAENARA"
misc_feature		complement(2646. .2669) /gene="SCI35.04c" /note="PS00017 ATP /GTP-binding site motif A (P-loop)"
gene		complement(2750. .5224) /gene="SCI35.05c" /note="SC01583"
CDS		complement(2750. .5224) /gene="SCI35.05c" /note="SCI35.05c, probable secreted arabinosidase, len: 824 aa; similar to e.g. TR_059218 (EMBL:U15178) arabinosidase (EC 3.2.1.55) (Bacteroides ovatus) (660 aa), fasta scores: opt: 1204 z-score: 1030.7 E(): 0, 41.4% identity in 601 aa overlap" /codon_start=1 /transl_table=1 /product="putative secreted arabinosidase"

NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penile & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: U.S.A.
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/309,512
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: 8145-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 854-3660
TELEFAX: (415) 854-3694
TELEX: 66141 PENNTE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 752 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acetobacter xylinum
US-08-309-512-8

Query Match 3.3%; Score 7; DB 1; Length 752;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEOSET 45
Db 304 AIEOSET 310

RESULT 8
PCT-US92-08756A-8
Sequence 8, Application PC/TUS9208756A
GENERAL INFORMATION:
APPLICANT: Tal, Ronny
APPLICANT: Benziman, Moshe
APPLICANT: Gelfand, David H.
APPLICANT: Ben-Bassat, Arie
APPLICANT: Calhoun, Roger D.
APPLICANT: Wong, Hing C.
TITLE OF INVENTION: CYCLIC DIGUANYLATE
TITLE OF INVENTION: METABOLIC ENZYMES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Limbach and Limbach
STREET: 2001 Ferry Building
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0,
SOFTWARE: Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08756A
FILING DATE: 19921014
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,218
FILING DATE: 29-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bortner, Scott R.
REGISTRATION NUMBER: 34,298
REFERENCE/DOCKET NUMBER: MEYR 20050 USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-433-4150
TELEFAX: 415-433-8716
TELEX: 278356
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 752 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acetobacter xylinum
PCT-US92-08756A-8

Query Match 3.3%; Score 7; DB 5; Length 752;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEOSET 45
Db 304 AIEOSET 310

RESULT 9
US-08-029-333-31
Sequence 31, Application US/08029333
Patent No. 539667
GENERAL INFORMATION:
APPLICANT: Frazier, William A.
APPLICANT: Kosfeld, Minh D.
TITLE OF INVENTION: Thrombospondin Receptor Binding Peptides
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SG
STREET: 800 N. Lindbergh Blvd.
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63167
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/029,333
FILING DATE: 19930305
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Meyer, Scott J.
REGISTRATION NUMBER: 25,275
REFERENCE/DOCKET NUMBER: 07-24(982)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)694-5435
TELEFAX: (314)694-5435
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: December 10, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
;
US-08-292-945-8

Query Match      3.3%; Score 7; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
Db 59 VAVSADP 65

RESULT 5
US-08-252-073A-8
; Sequence 8, Application US/08252073A
; Patent No. 5767073
; GENERAL INFORMATION:
; APPLICANT: Bing Lim, et al.
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center,
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30306-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,073A
; FILING DATE: 1-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
;

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
;
US-08-252-073A-8

Query Match      3.3%; Score 7; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
Db 59 VAVSADP 65

RESULT 6
PCT-US93-12074-8
; Sequence 8, Application PC/TUS9312074
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12074
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
;
PCT-US93-12074-8

Query Match      3.3%; Score 7; DB 5; Length 204;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
Db 59 VAVSADP 65

RESULT 7
US-08-309-512-8
; Sequence 8, Application US/08309512
; Patent No. 5759828
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Benzinan, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoun, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
;

```

Qy	61	HVGIVDFEGELAMRIEIRKGLKOKKROGDANVKEEGIVKAHLLIGVHDDIVSNEDYLAY	120
Db	61	HVGIVDFEGELAMRIEIRKGLKOKKROGDANVKEEGIVKAHLLIGVHDDIVSNEDYLAY	120
Qy	121	KLGDLPHTTHVVISDIDPQVVALSLEISDEGNIITWTSPFROFANVNHIGSLSTIDP IFG	180
Db	121	KLGDLPHTTHVVISDIDPQVVALSLEISDEGNIITWTSPFROFANVNHIGSLSTIDP IFG	180
Qy	181	VLSDLVTAIFODTVRKEMTKVLAPAFKRELEKN 213	
Db	181	VLSDLVTAIFODTVRKEMTKVLAPAFKRELEKN 213	

RESULT 2
US-09-46

```

; Sequence 2, Application US/08462778
; Patent No. 6077517
;
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kaw-Yan
; TITLE OF INVENTION: Allergenic Protein and Peptides from
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,778
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US9N 08/031,141
; FILING DATE: 12 March 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-462-778-2
;
; Query Match 11.7%; Score 25; DB 3; Length 215
; Best Local Similarity 100.0%; Pred. No. 1,5e-17;
; Matches 25; Conservative 0; Mismatches 0; Indels
;
; Oy 155 TSFEVROFANVYNNHIGGSIIDPIF 179
; |||||
; Db 155 TSFEVROFANVYNNHIGGSIIDPIF 179
;
; RESULT 3
; US-08-553-336A-2
; Sequence 2, Application US/08553336A
; Patent No. 6413738
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua

```

```

1  TITLE OF INVENTION: Allergenic Proteins and Peptides From
2  TITLE OF INVENTION: Allergenic Proteins and Peptides From
3  NUMBER OF SEQUENCES: 15
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: LAHIVE & COCKFIELD, LLP
6  STREET: 28 State Street
7  CITY: Boston
8  STATE: Massachusetts
9  COUNTRY: USA
10 ZIP: 02109
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.25
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/553,356A
20 FILING DATE: 10-JUN-1996
21 CLASSIFICATION: 424
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/081,540
24 FILING DATE: 22-JUNE-1993
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Jane E. Remillard
27 REGISTRATION NUMBER: 38,872
28 REFERENCE/DOCKET NUMBER: IMI-032CP2
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (617)227-7400
31 TELEFAX: (617)742-4214
32 INFORMATION FOR SEQ ID NO: 2:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 215 amino acids
35 TYPE: amino acid
36 TOPOLOGY: linear
37
38 MOLECULE TYPE: protein
39
40 US-08-353-336A-2

```

RESULT 4
 US-08-292-945-8
 Sequence 8, Application US/08292945
 Patent No. 5583478
 GENERAL INFORMATION:
 APPLICANT: Lim, Bing
 APPLICANT: Lelias, Jean-Michel
 APPLICANT: Adra, Chaker
 APPLICANT: Ko, Jone
 TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Kilpatrick & Cody
 STREET: 1100 Peachtree Street, Suite 2800
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: United States
 ZIP: 30309-4530
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/292,945
 FILING DATE:
 CLASSIFICATION: 435

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:14:25 ; Search time 15 Seconds

(without alignments)
417,805 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213
Sequence: 1 MMKFLIAAFAVAVSADPI.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213	100.0	213	4	US-08-553-336A-7
2	25	11.7	215	3	US-08-462-778-2
3	25	11.7	215	4	US-08-553-336A-2
4	7	3.3	204	1	US-08-292-945-8
5	7	3.3	204	1	US-08-252-073A-8
6	7	3.3	204	5	PCT-US93-12074-8
7	7	3.3	752	1	US-08-309-512-8
8	7	3.3	752	1	PCT-US92-08756A-8
9	6	2.8	10	4	US-08-029-333-31
10	6	2.8	10	4	US-08-391-820-31
11	6	2.8	12	1	US-07-616-910-54
12	6	2.8	12	1	US-08-405-200-2
13	6	2.8	12	1	US-08-405-200-15
14	6	2.8	12	5	PCT-US91-08497-54
15	6	2.8	18	1	US-07-895-252-10
16	6	2.8	18	1	US-08-072-283B-10
17	6	2.8	19	1	US-07-998-820-2
18	6	2.8	19	2	US-08-280-646-3
19	6	2.8	19	2	US-08-445-193-3
20	6	2.8	19	2	US-08-432-698-3
21	6	2.8	19	2	US-08-447-810-3
22	6	2.8	19	2	US-08-279-773-3
23	6	2.8	19	2	US-08-747-137-56
24	6	2.8	19	4	US-08-447-997-3
25	6	2.8	19	5	PCT-US95-09281-3
26	6	2.8	19	5	PCT-US95-09282-3
27	6	2.8	57	4	US-09-562-737-124

28	6	2.8	67	4	US-08-905-223-335	Sequence 335, App
29	6	2.8	102	4	US-09-383-586-34	Sequence 34, App1
30	6	2.8	115	6	5164490-6	Patent No. 5164490
31	6	2.8	116	4	US-09-562-737-125	Sequence 125, App
32	6	2.8	128	4	US-09-199-637A-150	Sequence 150, App
33	6	2.8	130	2	US-08-944-449-2	Sequence 2, App1
34	6	2.8	130	4	US-09-353-362-2	Sequence 2, App1
35	6	2.8	133	2	US-08-891-848-15	Sequence 15, App1
36	6	2.8	133	2	US-08-875-811-10	Sequence 10, App1
37	6	2.8	134	2	US-08-891-848-14	Sequence 14, App1
38	6	2.8	134	3	US-08-875-811-9	Sequence 9, App1
39	6	2.8	134	4	US-09-446-959-5	Sequence 5, App1
40	6	2.8	147	4	US-09-383-586-35	Sequence 35, App1
41	6	2.8	152	4	US-09-446-959-7	Sequence 7, App1
42	6	2.8	170	1	US-08-460-739-3	Sequence 3, App1
43	6	2.8	170	2	US-08-483-101-11	Sequence 11, App1
44	6	2.8	178	4	US-09-446-959-10	Sequence 10, App1
45	6	2.8	186	4	US-09-230-637-36	Sequence 36, App1

ALIGNMENTS

```
RESULT 1
US-08-553-336A-7
; Sequence 7, Application US/08553336A
; Patent No. 6413738
;
GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHYTE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)7227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-553-336A-7
;
Query Match 100.0%; Score 213; DB 4; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.6e-207; Indels 0; Gaps 0;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMKFLIAAFAVAVSADPIHYDKITEINKAIDDAIAIEQSEITDPMKVPDADKFER 60
Db 1 MMKFLIAAFAVAVSADPIHYDKITEINKAIDDAIAIEQSEITDPMKVPDADKFER 60
```

```

XX 05-MAY-1998; 98US-0073010.
PR 20-MAY-1997; 97US-0859381.
XX
XX (CORI-) CORIXA CORP.
XX
XX Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAM;
XX WPI; 1999-045314/04.
XX
PT Polypeptide comprising immunogenic Mycobacterium tuberculosis
PT antigen - useful for immunisation against M. tuberculosis infection
PT to treat or prevent tuberculosis, and in diagnosis of tuberculosis
XX
PS Example 1; Page 63; 100pp: English.
XX
CC This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans),
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induction) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC lack of sensitivity and specificity of existing diagnostic techniques.
CC
SQ Sequence 15 AA;

Query Match 2.8%; Score 6; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLT A 188
DB 3 SDVLT A 8

RESULT 50
AAW73826
ID AAW73826 standard; Peptide; 15 AA.
XX
XX AAW73826;
XX
XX 24-MAR-1999 (first entry)
XX
XX M. tuberculosis antigen clone ORF-2 peptide 2-23.
XX
XX Antigen: M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
XX infection.
XX
XX Mycobacterium tuberculosis.
XX
XX OS MO9853075-A2.
XX
XX PN 26-NOV-1998.
XX
XX PD 20-MAY-1998; 98WO-US10407.
XX
XX PF 05-MAY-1998; 98US-0073010.
XX
XX PR 20-MAY-1997; 97US-0859381.
XX
XX (CORI-) CORIXA CORP.

```

```

XX Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAM;
XX WPI; 1999-045314/04.
XX
XX Polypeptide comprising immunogenic Mycobacterium tuberculosis
XX antigen - useful for immunisation against M. tuberculosis infection
XX to treat or prevent tuberculosis, and in diagnosis of tuberculosis
XX
PS Example 1; Page 62; 100pp: English.
XX
CC This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans),
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induction) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC lack of sensitivity and specificity of existing diagnostic techniques.
CC
SQ Sequence 15 AA;

Query Match 2.8%; Score 6; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLT A 188
DB 8 SDVLT A 13

Search completed: February 21, 2003, 10:14:56
Job time : 39 secs

```

KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 XX infection.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9853076-A2.
 XX
 PD 26-NOV-1998;
 XX
 PF 20-MAY-1998; 98WO-US10514.
 XX
 PR 05-MAY-1998; 98US-0073009.
 PR 20-MAY-1997; 97US-0858998.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAM;
 DR WPI; 1999-045315/04.
 XX
 PT New isolated Mycobacterium tuberculosis antigens - used to develop
 PT products for the prevention, treatment and diagnosis of tuberculosis
 PT infection
 PS
 XX Example 1; Page 63; 104pp; English.
 CC This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans);
 CC e.g. they can be included with an acceptable carrier in pharmaceutical
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
 CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques.
 XX
 SQ Sequence 15 AA;
 Query Match 2.8%; Score 6; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 183 SDVLTA 188
 Db 8 SDVLTA 13
 RESULT 48
 AAW73718
 ID AAW73718 standard; Peptide; 15 AA.
 XX
 AC AAW73718;
 XX
 DT 24-MAR-1999 (first entry)
 XX
 DE M. tuberculosis antigen clone ORF-2 peptide 2-25.
 XX
 KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 XX infection.
 OS Mycobacterium tuberculosis.
 XX

PN WO9853076-A2.
 XX
 PD 26-NOV-1998.
 XX
 PF 20-MAY-1998; 98WO-US10514.
 XX
 PR 05-MAY-1998; 98US-0073009.
 PR 20-MAY-1997; 97US-0858998.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAM;
 DR WPI; 1999-045315/04.
 XX
 PT New isolated Mycobacterium tuberculosis antigens - used to develop
 PT products for the prevention, treatment and diagnosis of tuberculosis
 PT infection
 PS
 XX Example 1; Page 64; 104pp; English.
 CC This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans);
 CC e.g. they can be included with an acceptable carrier in pharmaceutical
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
 CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques.
 XX
 SQ Sequence 15 AA;
 Query Match 2.8%; Score 6; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 183 SDVLTA 188
 Db 3 SDVLTA 8
 RESULT 49
 AAW73828
 ID AAW73828 standard; Peptide; 15 AA.
 XX
 AC AAW73828;
 XX
 DT 24-MAR-1999 (first entry)
 XX
 DE M. tuberculosis antigen clone ORF-2 peptide 2-25.
 XX
 KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 XX infection.
 OS Mycobacterium tuberculosis.
 XX
 PN WO9853075-A2.
 XX
 PD 26-NOV-1998.
 XX
 PF 20-MAY-1998; 98WO-US10407.
 XX

CC AAR70302, AAR70303, and AAR70305 are synthetic peptides which are
 CC capable of binding the laminin binding domain. These peptides
 CC are useful in cell binding assays.

XX
 SQ Sequence 10 AA;

Query Match 2.8%; Score 6; DB 16; Length 10;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
 |||||

Db 4 VAVSAD 9

RESULT 45

ABG67586
 ID ABG67586 standard; Peptide; 10 AA.

XX
 AC ABG67586;

XX
 DT 07-OCT-2002 (first entry)

XX
 DE Human ADPI tryptic digest peptide #295.

XX
 KW Human; Alzheimer's disease; AD; brain tissue; ADF; ADPI;

KM Alzheimer's disease-associated feature; neuroprotective;
 KM Alzheimer's disease-associated protein isoform; nootropic;
 KM ADPI tryptic digest peptide.

XX
 OS Homo sapiens.

XX
 PN MO200246767-A2.

XX
 PD 13-JUN-2002.

XX
 PF 29-NOV-2001; 2001WO-GB05289.

XX
 PR 08-DEC-2000; 2000US-254431P.

XX
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX
 PI Herath HMAc, Parekh RB, Rohlf C;

XX
 DR WPI; 2002-508575/54.

XX
 PT Screening, diagnosis or prognosis of Alzheimer's disease in subject,

PT comprises detecting Alzheimer's disease-associated features or
 PT Alzheimer's disease-associated protein isoforms in brain tissue
 PT from the subject

XX
 PS Claim 7; Page 66; 427pp; English.

XX The present invention relates to methods and compositions for the
 CC screening, diagnosis or prognosis of Alzheimer's disease (AD) in
 CC a subject. The method comprises analysing a sample of brain tissue
 CC from a subject by 2D electrophoresis to generate a 2D array of
 CC Alzheimer's disease-associated features (ADFs), whose relative
 CC abundance correlates with the presence, absence, stage or severity of
 CC AD and comparing the abundance of each feature with the abundance of
 CC that chosen feature in brain tissue from persons free from AD. The
 CC invention also describes Alzheimer's disease-associated protein
 CC isoforms (ADPIs) detectable in brain tissue. The methods and
 CC compositions of the invention are useful for the screening, diagnosis
 CC or prognosis of AD in a subject, for determining the stage or severity
 CC of AD in a subject, for identifying a subject at risk of developing AD,
 CC or for monitoring the effect of therapy administered to a subject
 CC having AD. Antibodies capable of binding to ADPIs are useful for
 CC treating or preventing AD, and for determining the efficacy of a given
 CC treatment regime. An agent that modulates the activity of ADPI is
 CC useful in the manufacture of a medicament for the treatment or
 CC prevention of AD in a subject. ABG67292-ABG68038 represent human ADPI
 CC tryptic digest peptides.

XX
 SQ Sequence 10 AA;

Query Match 2.8%; Score 6; DB 23; Length 10;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 HLLIGV 107
 |||||

Db 1 HLLIGV 6

RESULT 46

AAW69442
 ID AAW69442 standard; Protein; 11 AA.

XX
 AC AAW69442;

XX
 DT 10-DEC-1998 (first entry)

XX
 DE Rat phospholipase A1 fragment.

XX
 KW Phospholipase A1; phosphatidyl-serine hydrolysis; serine phospholipid;

XX
 PLAI; enzyme; rat.

XX
 OS Rattus sp.

XX
 FH Key Location/Qualifiers

FT Misc-difference 10 /note="unspecified amino acid"

XX
 PN JP10201479-A.

XX
 PD 04-AUG-1998.

XX
 PF 23-JAN-1997; 97JP-0024269.

XX
 PR 23-JAN-1997; 97JP-0024269.

XX
 PA (TORA) TORAY IND INC.

XX
 DR WPI; 1998-474493/41.

XX
 PT Phospholipase A1 and nucleic acid coding for it - having substrate
 PT specificity against serine phospholipid

XX
 PS Example 1; Page 5; 13pp; Japanese.

XX
 CC This sequence is a fragment of the rat phospholipase A1 (PLA1) of the
 CC invention. PLA1 is capable of hydrolysing phosphatidyl-serine, and has
 CC substrate specificity for serine phospholipids.

XX
 SQ Sequence 11 AA;

Query Match 2.8%; Score 6; DB 19; Length 11;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LSLIEIS 147
 |||||

Db 1 LSLIEIS 6

RESULT 47

AAW73716
 ID AAW73716 standard; Peptide; 15 AA.

XX
 AC AAW73716;

XX
 DT 24-MAR-1999 (first entry)

XX
 DE M. tuberculosis antigen clone ORF-2 peptide 2-23.

XX OS Drosophila melanogaster.
 XX PN WO2001171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PMD, Myers EM;
 XX DR WPI; 2001-656860/75.
 XX DR N-PSDB; ABL06822.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX PS Disclosure; SEQ ID NO 14949; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB857737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC SQ Sequence 2836 AA;
 QY Query Match 3.3%; Score 7; DB 22; Length 2836;
 Best Local Similarity 100.0%; Pred. No. 1e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 72 AMRNEA 78
 1376 AMRNEA 1382
 RESULT 43
 AAB23790
 ID AAB23790 standard; peptide; 8 AA.
 XX AC AAB23790;
 XX DT 12-JAN-2001 (first entry)
 XX DE Voltage sensitive cell adhesion molecule synthetic peptide PA22-2.
 XX KW Voltage sensitive cell; VSC; adhesion molecule; binding molecule;
 KW neuron; neurite promoting domain; hybrid electrical device;
 KW electrical junction; transistor; artificial chemical synapse.
 XX OS Synthetic.
 XX PN WO200051191-A1.
 XX PD 31-AUG-2000.
 XX PF 22-FEB-2000; 2000WO-IL00112.
 XX PR 22-FEB-1999; 99US-0121237.
 XX PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

XX PI Yitzchaik S, Shappir J, Spira ME;
 XX DR WPI; 2001-015536/02.
 XX PT A hybrid electrical device with biological components that provide an
 XX artificial chemical synapse -
 XX PS Example 6; Page 13; 48pp; English.
 XX CC The present invention describes a hybrid electrical device with
 CC biological components that provide an artificial chemical synapse.
 CC An electrical junction (I) between one transistor and at least 1
 CC voltage-sensitive cell (VSC) characterised by at least 1 of the features
 CC selected from: (i) voltage transfer between the transistor and the VSC
 CC is bidirectional; (ii) there is no DC bias between the transistor and
 CC the solution containing the VSC; (iii) the VSC is anchored to the
 CC external surface of the transistor by binding moieties, optionally
 CC through spacers; and/or (iv) the voltage transfer between the membra
 CC of the VSC and the external surface of the transistor, and between the
 CC external surface of the transistor and the membrane of the VSC is
 CC mediated by hyper-polarisable chromophores. As an artificial chemical
 CC synapse i.e., a junction between a cell (which secretes an agent) and a
 CC transistor bearing receptors for the agent. Binding of the agent to the
 CC receptor changes an electrical property of the transistor. The present
 CC sequence represents a synthetic peptide, designated PA 22-2, which can
 CC be used as a VSC adhesion molecule, where the VSC is a neuron, which is
 CC used in an example from the present invention.
 XX SQ Sequence 8 AA;
 QY Query Match 2.8%; Score 6; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 13 VAVSAD 18
 2 VAVSAD 7
 RESULT 44
 AAR70305
 ID AAR70305 standard; peptide; 10 AA.
 XX AC AAR70305;
 XX DT 14-NOV-1995 (first entry)
 XX DE Laminin (LM) binding peptide, LM-2.
 XX KW Laminin; LM; soluble; receptor; cell-adhesion.
 XX OS Synthetic.
 XX PN US5399667-A.
 XX PD 21-MAR-1995.
 XX PF 05-MAR-1993; 93US-0029333.
 XX PR 05-MAR-1993; 93US-0029333.
 XX PA (UNIW) UNIV WASHINGTON.
 XX PI Frazier WA, Kosfeld MD;
 XX DR WPI; 1995-130736/17.
 XX PT Thrombospondin receptor binding peptide(s) - comprise tri-peptide
 PT sequence Val-Val-Met (VVM)
 XX PS Disclosure; Figure 4; 25pp; English.
 XX

```

RESULT 40
AAR38153
ID AAR38153 standard; protein; 752 AA.
XX
AC AAR38153;
XX
DT 13-OCT-1993 (first entry)
XX
DE Acetobacter diacyanolate phosphodiesterase PDEA2.
XX
KW Cyclic diacyanolate; diacyanolate phosphodiesterase;
XX diacyanolate cyclase; cellulose production; cdg2 operon.
XX
OS Acetobacter xylinum.
XX
FH Key Location/Qualifiers
FT MISC-difference 77 /note= "Val deduced from GG"
FT MISC-difference 212 /note= "Met deduced from TAG"
FT MISC-difference 215 /note= "Arg deduced from GC"
FT MISC-difference 232 /note= "Met deduced from ATC"
FT MISC-difference 233 /note= "Leu deduced from GTG"
FT MISC-difference 247 /note= "Gly deduced from GC"
FT MISC-difference 271 /note= "Gln deduced from CAGC"
FT MISC-difference 275 /note= "Val deduced from GCTG"
FT MISC-difference 388 /note= "Asp deduced from GAA"
FT MISC-difference 538 /note= "Ala deduced from CC"
XX
PN WO9311244-A.
XX
PD 10-JUN-1993.
XX
PF 14-OCT-1992; 92WO-US08756.
XX
PR 29-NOV-1991; 91US-0800218.
XX
PA (WEYE ) WEYERRHAUSER CO.
XX
PI Ben-Bassat A, Benzman M, Calhoon RD, Gelfand DH;
PI Tal R, Wong HC;
XX
DR WPI: 1993-197062/24.
DR N-PSDB; AAQ43661.
XX
XX Polynucleotide sequence from Acetobacter cdg operon - encodes
XX cyclic di:guanosine mono:phosphate degradation enzymes e.g.
XX 3-phosphodiesterase isozyme
XX
PS Claim 5; Page 80-83; 98pp; English.
XX
CC The amino acid sequence of protein PDEA2 was deduced from the 1st.
CC open reading frame of the cdg2 operon. The protein is a diacyanolate
CC phosphodiesterase A', i.e. it enzymatically cleaves a single
CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG.
CC See also AAR38154.
XX
SQ Sequence 752 AA:

```

```

Query Match 3.3%; Score 7; DB 14; Length 752;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 39 AIEOSET 45
      |||||
Db 304 AIEOSET 310

RESULT 41
ABB93713
ID ABB93713 standard; protein; 791 AA.
XX
AC ABB93713;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2924.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
OS WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidner M;
XX
DR WPI: 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 2924; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 791 AA:

```

```

Query Match 3.3%; Score 7; DB 23; Length 791;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 AIDDATA 38
      |||||
Db 299 AIDDATA 305

RESULT 42
ABB62719
ID ABB62719 standard; protein; 2836 AA.
XX
AC ABB62719;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 14949.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

```

CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

XX
SQ Sequence 474 AA;

Query Match 3.3%; Score 7; DB 23; Length 474;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 HVISDIQ 136
DB 440 HVISDIQ 446

RESULT 38

ABB62629 ID ABB62629 standard; Protein; 511 AA.

XX ABB62629;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 14679.

DE Drosophila melanogaster polypeptide SEQ ID NO 14679.

KW Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2001; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR MPI: 2001-656860/75.

DR N-PSDB; ABL06732.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure: SEQ ID NO 14679; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 511 AA;

Query Match 3.3%; Score 7; DB 22; Length 511;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 KYLAPAF 206

DB 129 KYLAPAF 135

RESULT 39

ABG25418 ID ABG25418 standard; Protein; 581 AA.

XX ABG25418;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #25409.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR MPI: 2001-639362/73.

XX N-PSDB; AAS89605.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID NO 55777; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 581 AA;

Query Match 3.3%; Score 7; DB 22; Length 581;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDALIAI 40
DB 289 DDALIAI 295

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX Lactococcus lactis ILL403.
 OS
 XX
 PN FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Bolotine A, Sorokline A, Renault P, Ehrlich SD;
 XX
 DR WPI; 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification or Lactococcus
 PT lactis and related species -
 PS
 PS Claim 6: SEQ ID No 1166; 2504bp; French.
 XX
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SQ Sequence 434 AA;
 XX
 QY 5 LLIATAVA 11
 IIIIII
 Db 372 LLIATAVA 378
 XX
 RESULT 36
 ABB63535
 ID ABB63535 standard; Protein; 460 AA.
 XX
 AC ABB63535;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 17397.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX

PI Venter JC, Adams M, Li PWD, Myers EM;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL07638.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 PS Disclosure; SEQ ID NO 17397; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
 CC sequences (ABL01840-ABL16175).
 CC (ABAB57737-ABAB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SQ Sequence 460 AA;
 XX
 QY 3 KFLIIAA 9
 IIIIII
 Db 326 KFLIIAA 332
 XX
 RESULT 37
 ABB92783
 ID ABB92783 standard; Protein; 474 AA.
 XX
 AC ABB92783;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 DE Herbicidally active polypeptide SEQ ID NO 1994.
 XX
 KM Herbicidal; plant; agriculture; herbicide.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Tietjen K, Weidler M;
 XX
 DR WPI; 2002-269010/31.
 XX
 PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 PS
 PS Claim 5; SEQ ID NO 1994; 261bp + Sequence Listing; English.
 XX
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using

CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.

XX SQ Sequence 300 AA;

Query Match 3.3%; Score 7; DB 21; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 KRELEKN 213
 DB 214 KRELEKN 220

RESULT 33
 AAR3439
 ID AAR3439 standard; Protein; 354 AA.

XX AC AAR3439;

XX DT 17-DEC-2001 (updated)
 XX DT 06-JUL-1993 (first entry)

XX DE Ornithine cyclodeaminase C58 from Ti plasmid pTIC58.

XX KW mu-crystallins; drug targeting; nervous acting drugs; CNS; neural;
 KW neuronal; neurotransmitter agents; neuromuscular agents; NMJ;
 KW neuromuscular junctions; memory agents; Alzheimers disease;
 KW CNS depressants; CNS stimulants; tranquilizers; muscle relaxants;
 KW antispasmodics; analgesics; anesthetics; anticonvulsants;
 KW antiepileptic agents; antianxiety agents; hallucinogens; sedatives;
 KW hypnotics.

XX OS Agrobacterium tumefaciens.

XX PN USN7844304-N.

XX PD 01-JAN-1993.

XX PF 28-FEB-1992; 92US-0844304.

XX PR 28-FEB-1992; 92US-0844304.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

XX PI Kim R, Wistow G;

XX DR WPI; 1993-093573/11.

XX PT New mu-crystalline proteins - having ornithine cyclo-deaminase
 PT activity, used in diagnosis and treatment of disorders in
 XX ornithine metabolism

XX PS Disclosure; Page 34; 60pp; English.

XX CC This sequence represents ornithine cyclodeaminase (OCD) from
 CC Agrobacterium Ti plasmid pTIC58. It shows approximately 30%
 CC homology with the kangaroo eye lens protein mu-crystallin.
 CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NTIS applications to prevent clashes with ongoing US
 CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
 XX

XX SQ Sequence 354 AA;

Query Match 3.3%; Score 7; DB 14; Length 354;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 FGVLSDV 185

DB 96 FGVLSDV 102

RESULT 34
 AAR3440
 ID AAR3440 standard; Protein; 357 AA.

XX AC AAR3440;

XX DT 17-DEC-2001 (updated)
 XX DT 06-JUL-1993 (first entry)

XX DE Ornithine cyclodeaminase Ach5 from Ti plasmid Ach5.

XX KW mu-crystallins; drug targeting; nervous acting drugs; CNS; neural;
 KW neuronal; neurotransmitter agents; neuromuscular agents; NMJ;
 KW neuromuscular junctions; memory agents; Alzheimers disease;
 KW CNS depressants; CNS stimulants; tranquilizers; muscle relaxants;
 KW antispasmodics; analgesics; anesthetics; anticonvulsants;
 KW antiepileptic agents; antianxiety agents; hallucinogens; sedatives;
 KW hypnotics.

XX OS Agrobacterium tumefaciens.

XX PN USN7844304-N.

XX PD 01-JAN-1993.

XX PF 28-FEB-1992; 92US-0844304.

XX PR 28-FEB-1992; 92US-0844304.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

XX PI Kim R, Wistow G;

XX DR WPI; 1993-093573/11.

XX PT New mu-crystalline proteins - having ornithine cyclo-deaminase
 PT activity, used in diagnosis and treatment of disorders in
 XX ornithine metabolism

XX PS Disclosure; Page 34; 60pp; English.

XX CC This sequence represents ornithine cyclodeaminase (OCD) from
 CC Agrobacterium Ti plasmid Ach5. It shows approximately 30%
 CC homology with the kangaroo eye lens protein mu-crystallin.
 CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NTIS applications to prevent clashes with ongoing US
 CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
 XX

XX SQ Sequence 357 AA;

Query Match 3.3%; Score 7; DB 14; Length 357;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 FGVLSDV 185
 DB 96 FGVLSDV 102

RESULT 35
 ABB54464
 ID ABB54464 standard; Protein; 434 AA.

XX AC ABB54464;

XX DT 16-MAY-2002 (first entry)

XX DE Lactococcus lactis protein pbux.

```

PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

```

```

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.3%; Score 7; DB 21; Length 215;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 ULTRAHQ 191
Db 119 ULTRAHQ 125

RESULT 32
AAB18216
ID AAB18216 standard; Protein; 300 AA.
XX
AC AAB18216;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:73.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (GARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI: 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX
PS Disclosure; Page 171-172; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of proteins encoded by it will help to expand
CC subsequent identification of proteins encoded by chromosome 2 and the
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito

```

PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.3%; Score 7; DB 21; Length 214;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLFATFQ 191
Db 117 VLFATFQ 123

RESULT 31
AGS55906
ID AGS55906 standard; Protein; 215 AA.
XX
AC AGS55906;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 71760.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144362.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.

PR	05-MAY-1999;	990S-0132485.
PR	06-MAY-1999;	990S-0132486.
PR	06-MAY-1999;	990S-0132487.
PR	07-MAY-1999;	990S-0132863.
PR	11-MAY-1999;	990S-0134256.
PR	14-MAY-1999;	990S-0134218.
PR	14-MAY-1999;	990S-0134219.
PR	14-MAY-1999;	990S-0134221.
PR	18-MAY-1999;	990S-0134370.
PR	18-MAY-1999;	990S-0134768.
PR	19-MAY-1999;	990S-0134941.
PR	20-MAY-1999;	990S-0135124.
PR	21-MAY-1999;	990S-0135353.
PR	24-MAY-1999;	990S-0135629.
PR	25-MAY-1999;	990S-0136021.
PR	27-MAY-1999;	990S-0136392.
PR	28-MAY-1999;	990S-0136782.
PR	01-JUN-1999;	990S-0137222.
PR	03-JUN-1999;	990S-0137528.
PR	04-JUN-1999;	990S-0137502.
PR	07-JUN-1999;	990S-0137724.
PR	08-JUN-1999;	990S-0138094.
PR	10-JUN-1999;	990S-0138540.
PR	10-JUN-1999;	990S-0138847.
PR	14-JUN-1999;	990S-0139119.
PR	16-JUN-1999;	990S-0139452.
PR	16-JUN-1999;	990S-0139453.
PR	17-JUN-1999;	990S-0139492.
PR	18-JUN-1999;	990S-0139454.
PR	18-JUN-1999;	990S-0139455.
PR	18-JUN-1999;	990S-0139456.
PR	18-JUN-1999;	990S-0139457.
PR	18-JUN-1999;	990S-0139458.
PR	18-JUN-1999;	990S-0139459.
PR	18-JUN-1999;	990S-0139460.
PR	18-JUN-1999;	990S-0139461.
PR	18-JUN-1999;	990S-0139462.
PR	18-JUN-1999;	990S-0139463.
PR	18-JUN-1999;	990S-0139750.
PR	18-JUN-1999;	990S-0139763.
PR	21-JUN-1999;	990S-0139817.
PR	22-JUN-1999;	990S-0139899.
PR	23-JUN-1999;	990S-0140353.
PR	23-JUN-1999;	990S-0140354.
PR	24-JUN-1999;	990S-0140695.
PR	28-JUN-1999;	990S-0140823.
PR	29-JUN-1999;	990S-0140991.
PR	30-JUN-1999;	990S-0141287.
PR	01-JUL-1999;	990S-0141842.
PR	02-JUL-1999;	990S-0142055.
PR	06-JUL-1999;	990S-0142390.
PR	08-JUL-1999;	990S-0142803.
PR	09-JUL-1999;	990S-0142920.
PR	12-JUL-1999;	990S-0142977.
PR	13-JUL-1999;	990S-0143542.
PR	14-JUL-1999;	990S-0143624.
PR	15-JUL-1999;	990S-0144005.
PR	16-JUL-1999;	990S-0144085.
PR	16-JUL-1999;	990S-0144086.
PR	19-JUL-1999;	990S-0144325.
PR	19-JUL-1999;	990S-0144331.
PR	19-JUL-1999;	990S-0144332.
PR	19-JUL-1999;	990S-0144333.
PR	19-JUL-1999;	990S-0144334.
PR	19-JUL-1999;	990S-0144335.
PR	20-JUL-1999;	990S-0144632.
PR	20-JUL-1999;	990S-0144632.
PR	20-JUL-1999;	990S-0144684.
PR	21-JUL-1999;	990S-0144814.
PR	21-JUL-1999;	990S-0145086.
PR	21-JUL-1999;	990S-0145088.
PR	22-JUL-1999;	990S-0145085.
<hr/>		
PR	22-JUL-1999;	990S-0145087.
PR	22-JUL-1999;	990S-0145089.
PR	22-JUL-1999;	990S-0145192.
PR	23-JUL-1999;	990S-0145145.
PR	23-JUL-1999;	990S-0145218.
PR	23-JUL-1999;	990S-0145224.
PR	26-JUL-1999;	990S-0145276.
PR	27-JUL-1999;	990S-0145913.
PR	27-JUL-1999;	990S-0145918.
PR	27-JUL-1999;	990S-0145919.
PR	28-JUL-1999;	990S-0145951.
PR	02-AUG-1999;	990S-0146386.
PR	02-AUG-1999;	990S-0146388.
PR	02-AUG-1999;	990S-0146389.
PR	03-AUG-1999;	990S-0147038.
PR	04-AUG-1999;	990S-0147204.
PR	04-AUG-1999;	990S-0147302.
PR	05-AUG-1999;	990S-0147192.
PR	05-AUG-1999;	990S-0147260.
PR	06-AUG-1999;	990S-0147303.
PR	06-AUG-1999;	990S-0147416.
PR	09-AUG-1999;	990S-0147493.
PR	09-AUG-1999;	990S-0147935.
PR	10-AUG-1999;	990S-0148171.
PR	11-AUG-1999;	990S-0148319.
PR	12-AUG-1999;	990S-0148341.
PR	13-AUG-1999;	990S-0148565.
PR	13-AUG-1999;	990S-0148684.
PR	16-AUG-1999;	990S-0149368.
PR	17-AUG-1999;	990S-0149175.
PR	18-AUG-1999;	990S-0149426.
PR	20-AUG-1999;	990S-0149722.
PR	20-AUG-1999;	990S-0149723.
PR	20-AUG-1999;	990S-0149929.
PR	23-AUG-1999;	990S-0149902.
PR	23-AUG-1999;	990S-0149930.
PR	25-AUG-1999;	990S-0150566.
PR	26-AUG-1999;	990S-0150884.
PR	27-AUG-1999;	990S-0151065.
PR	27-AUG-1999;	990S-0151066.
PR	27-AUG-1999;	990S-0151080.
PR	30-AUG-1999;	990S-0151303.
PR	31-AUG-1999;	990S-0151438.
PR	01-SEP-1999;	990S-0151930.
PR	07-SEP-1999;	990S-0152363.
PR	10-SEP-1999;	990S-0153070.
PR	13-SEP-1999;	990S-0153758.
PR	16-SEP-1999;	990S-0154039.
PR	16-SEP-1999;	990S-0154039.
PR	20-SEP-1999;	990S-0154779.
PR	22-SEP-1999;	990S-0155139.
PR	23-SEP-1999;	990S-0155486.
PR	24-SEP-1999;	990S-0155659.
PR	28-SEP-1999;	990S-0156596.
PR	29-SEP-1999;	990S-0156596.
PR	04-OCT-1999;	990S-0157117.
PR	05-OCT-1999;	990S-0157753.
PR	06-OCT-1999;	990S-0157865.
PR	07-OCT-1999;	990S-0158029.
PR	08-OCT-1999;	990S-0158232.
PR	12-OCT-1999;	990S-0158369.
PR	13-OCT-1999;	990S-0159293.
PR	13-OCT-1999;	990S-0159294.
PR	13-OCT-1999;	990S-0159295.
PR	14-OCT-1999;	990S-0159329.
PR	14-OCT-1999;	990S-0159330.
PR	14-OCT-1999;	990S-0159331.
PR	14-OCT-1999;	990S-0159637.
PR	14-OCT-1999;	990S-0159638.
PR	18-OCT-1999;	990S-0159584.
PR	21-OCT-1999;	990S-0160741.
PR	21-OCT-1999;	990S-0160767.
PR	21-OCT-1999;	990S-0160768.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.3%; Score 7; DB 21; length 205;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 VLTAIFQ 191
|||||||
Db 107 VLTAIFQ 113

RESULT 30
AAG5323
ID AAG5323 standard; Protein; 214 AA.
XX
AC AAG5323;
XX
DT 18-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 70919.
XX
KM Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.

PN W09413802-A.
XX
PD 23-JUN-1994.
XX
PF 10-DEC-1993; 93MO-US12074.
XX
PR 10-DEC-1992; 92US-0990337.
XX
PA (BETH-) BETH ISRAEL HOSPITAL ASSOC.
XX (CYTO-) CYTOMED INC.
XX
PI Adra CN, Ko JL, Lelias J, Lim B;
XX
DR WPI; 1994-217878/26.
XX
PT New D4 protein preferentially expressed in haematopoietic cells -
PT for use as a diagnostic, in research and for treatment of
PT Leukemia tumours, and in bone marrow transplants.
XX
PS Disclosure; Page 56-57; 76pp; English.
XX
CC Differential screening of a subtractive hematopoietic cDNA
CC library yielded D4 cDNA, which encoded a protein that showed
CC GTP-dissociation inhibitor activity. The cDNA sequence was
CC determined (AA068715) and the amino acid sequence corresponding
CC to the longest ORF was deduced (AAR54112). Equivalent sequences
CC for mouse D4 are given in AA068716 and AAR54113. The cattle D4
CC protein sequence was also provided (AAR54114).
XX
SQ Sequence 204 AA;

Query Match 3.3%; Score 7; DB 15; Length 204;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
Db 59 VAVSADP 65

RESULT 29
AAG59753
ID AAG59753 standard; Protein: 205 AA.
AC AAG59753;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 77322.
XX
KM Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155148.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158362.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 3.3%; Score 7; DB 21; Length 185;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAFQ 191
 DB 88 VLTAFQ 94

RESULT 27
 ABB06132
 ID ABB06132 standard; Protein: 193 AA.

XX ABB06132:
 AC
 XX
 DT 10-MAY-2002 (first entry)
 XX
 DE Human NS protein sequence SEQ ID NO:224.
 XX

KW Human; cytosolic; osteopathic; gynaecological; neuroprotective;
 KW antirheumatic; antiarthritic; antipsoriatic; ophthalmological; anti-HIV;
 KW vasotropic; antiarteriosclerotic; antiinflammatory; dermatological;
 KW anorectic; muscular; antinfertility; cardiovascular; anticoagulant;
 KW antilibrinolytic; hypotension; antisthmatic; immunomodulator; cardiac;
 KW anticonvulsant; antidiabetic; tranquilliser; antidepressant; neuroleptic;
 KW gastrointestinal; virocidic; antitumor; cerebroprotective; nootropic;
 KW contraceptive; vaccine; gene therapy; cancer; osteoporosis; dystonia;
 KW endometriosis; degenerative disease; multiple sclerosis; psoriasis;
 KW rheumatoid arthritis; cataract; restenosis; atherosclerosis; glaucoma;
 KW inflammation; skin disorder; obesity; muscular dystrophy; AIDS;
 KW infertility; cardiovascular disease; coagulation disease; hypertension;
 KW ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;

KW diabetes; anxiety; depression; schizophrenia; viral disease; stroke;
 KW gastric ulcer; Alzheimer's disease.
 XX
 OS Homo sapiens.
 XX
 PN W0200206315-A2.
 XX
 PD 24-JAN-2002.
 XX
 PE 17-JUL-2001; 2001WO-1100653.
 XX
 PR 18-JUL-2000; 2000IL-0137345.
 PR 15-DEC-2000; 2000IL-0140354.
 XX
 PA (COMP-) COMPUGEN LTD.
 XX
 PI Mintz L, Freilich S, Bernstein J;
 XX
 DR WPI; 2002-155037/20.
 DR N-PSDB; ABL39786.
 XX
 PT One hundred and twenty eight novel nucleic acid sequences, useful for
 PT treating and diagnosing e.g. cancer, asthma and Alzheimer's -
 PS Claim 6; Page 260-261; 290pp; English.

CC ABL39691 to ABL39818 represent novel human nucleic acid sequences
 CC encoding the proteins given in ABB06037 to ABB06164. The novel sequences
 CC (NS) can have cytosolic, osteopathic, gynaecological, neuroprotective,
 CC antirheumatic, antiarthritic, antipsoriatic, ophthalmological, virocidic,
 CC vasotropic, antiarteriosclerotic, antiinflammatory, dermatological,
 CC anorectic, muscular, anti-HIV, antinfertility, cardiovascular,
 CC anticoagulant, antifibrinolytic, hypotension, antisthmatic, cardiac,
 CC immunomodulator, anticonvulsant, antidiabetic, tranquilliser, antitumor,
 CC antidepressant, gastrointestinal, aeturoleptic, cerebroprotective,
 CC nootropic and contraceptive activities. The NS can be used in vaccines,
 CC gene therapy and antisense therapy. Nucleic acids, expression vectors and
 CC antibodies from the present invention can be used for treating and
 CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative
 CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
 CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
 CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
 CC disease, coagulation disease, ischaemia, hypertension, asthma, immune
 CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,
 CC depression, schizophrenia, viral disease, gastric ulcers, stroke,
 CC Alzheimer's disease and as a contraceptive.

XX Sequence 193 AA;

Query Match 3.3%; Score 7; DB 23; Length 193;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
 DB 59 VAVSADP 65

RESULT 28
 AAR54114
 ID AAR54114 standard; Protein: 204 AA.

XX AAR54114:
 AC
 XX
 DT 20-JAN-1995 (first entry)
 XX
 DE GDI D4 protein.
 XX

KW D4, GTP-dissociation inhibitor; GDI; hematopoietic cell; diagnostic;
 KW leukemia; tumor bone marrow transplant.
 XX
 OS Bos taurus.
 XX

XX	EP1033405-A2.
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128717.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130440.
PR	23-APR-1999; 99US-0130510.
PR	28-APR-1999; 99US-0130891.
PR	30-APR-1999; 99US-0131449.
PR	04-MAY-1999; 99US-0132407.
PR	05-MAY-1999; 99US-0132484.
PR	06-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	07-MAY-1999; 99US-0132487.
PR	11-MAY-1999; 99US-0132863.
PR	14-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134994.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135129.
PR	24-MAY-1999; 99US-0135353.
PR	25-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138540.
PR	14-JUN-1999; 99US-0138847.
PR	16-JUN-1999; 99US-0139453.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139899.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140695.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUN-1999; 99US-0141287.
PR	01-JUL-1999; 99US-0141842.
PR	01-JUL-1999; 99US-0142354.
PR	02-JUL-1999; 99US-0142053.
PR	06-JUL-1999; 99US-0142390.
PR	08-JUL-1999; 99US-0142803.
PR	09-JUL-1999; 99US-0142920.
PR	12-JUL-1999; 99US-0142977.
PR	13-JUL-1999; 99US-0143542.
PR	14-JUL-1999; 99US-0143624.
PR	15-JUL-1999; 99US-0144005.
PR	16-JUL-1999; 99US-0144085.
PR	16-JUL-1999; 99US-0144086.
PR	19-JUL-1999; 99US-0144332.
PR	19-JUL-1999; 99US-0144331.
PR	19-JUL-1999; 99US-0144332.
PR	19-JUL-1999; 99US-0144334.
PR	19-JUL-1999; 99US-0144334.
PR	20-JUL-1999; 99US-0144335.
PR	20-JUL-1999; 99US-0144632.
PR	20-JUL-1999; 99US-0144884.
PR	21-JUL-1999; 99US-0144814.
PR	21-JUL-1999; 99US-0145086.
PR	21-JUL-1999; 99US-0145088.
PR	22-JUL-1999; 99US-0145085.
PR	22-JUL-1999; 99US-0145087.
PR	22-JUL-1999; 99US-0145089.
PR	22-JUL-1999; 99US-0145109.
PR	22-JUL-1999; 99US-0145119.
PR	23-JUL-1999; 99US-0145145.
PR	23-JUL-1999; 99US-0145145.
PR	23-JUL-1999; 99US-0145145.
PR	23-JUL-1999; 99US-0145224.
PR	26-JUL-1999; 99US-0145276.
PR	27-JUL-1999; 99US-0145913.
PR	27-JUL-1999; 99US-0145918.
PR	27-JUL-1999; 99US-0145919.
PR	28-JUL-1999; 99US-0145951.
PR	02-AUG-1999; 99US-0146386.
PR	02-AUG-1999; 99US-0146388.
PR	03-AUG-1999; 99US-0146389.
PR	

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142820.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.3%; Score 7; DB 21; Length 183;

Best Local Similarity 100.0%; Pred. No. 88; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAFQ 191

DB 85 VLTAFQ 91

RESULT 26

AAAG5324

ID AAAG5324 standard; Protein; 185 AA.

XX

AC AAAG5324;

XX 18-OCT-2000 (first entry)

DT

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 70920.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 13-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 3.3%; Score 7; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 VLTALFQ 191
 |||||
 Db 87 VLTALFQ 93

RESULT 25
 AAG59754

ID AAG59754 standard; Protein; 183 AA.

AC AAG59754;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 77323.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130077.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132047.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134321.
 PR 18-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 21-JUN-1999; 99US-0139763.
 PR 22-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.

ID AAG55907 standard; Protein; 183 AA.
XX
AC AAG55907;
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 71761.
DE
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130049.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.

PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145513.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 06-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.

PR 18-JUN-1999; 9905-0139454.
 PR 18-JUN-1999; 9905-0139455.
 PR 18-JUN-1999; 9905-0139456.
 PR 18-JUN-1999; 9905-0139457.
 PR 18-JUN-1999; 9905-0139458.
 PR 18-JUN-1999; 9905-0139459.
 PR 18-JUN-1999; 9905-0139460.
 PR 18-JUN-1999; 9905-0139461.
 PR 18-JUN-1999; 9905-0139462.
 PR 18-JUN-1999; 9905-0139463.
 PR 18-JUN-1999; 9905-0139750.
 PR 18-JUN-1999; 9905-0139763.
 PR 21-JUN-1999; 9905-0139817.
 PR 22-JUN-1999; 9905-0139899.
 PR 23-JUN-1999; 9905-0140353.
 PR 23-JUN-1999; 9905-0140354.
 PR 24-JUN-1999; 9905-0140695.
 PR 28-JUN-1999; 9905-0140823.
 PR 29-JUN-1999; 9905-0140991.
 PR 30-JUN-1999; 9905-0141287.
 PR 01-JUL-1999; 9905-0141842.
 PR 01-JUL-1999; 9905-0142154.
 PR 02-JUL-1999; 9905-0142055.
 PR 06-JUL-1999; 9905-0142390.
 PR 08-JUL-1999; 9905-0142803.
 PR 09-JUL-1999; 9905-0142920.
 PR 12-JUL-1999; 9905-0142977.
 PR 13-JUL-1999; 9905-0143542.
 PR 14-JUL-1999; 9905-0143624.
 PR 15-JUL-1999; 9905-0144005.
 PR 16-JUL-1999; 9905-0144086.
 PR 16-JUL-1999; 9905-0144086.
 PR 19-JUL-1999; 9905-0144325.
 PR 19-JUL-1999; 9905-0144331.
 PR 19-JUL-1999; 9905-0144332.
 PR 19-JUL-1999; 9905-0144333.
 PR 19-JUL-1999; 9905-0144334.
 PR 19-JUL-1999; 9905-0144335.
 PR 20-JUL-1999; 9905-0144352.
 PR 20-JUL-1999; 9905-0144632.
 PR 20-JUL-1999; 9905-0144884.
 PR 21-JUL-1999; 9905-0144814.
 PR 21-JUL-1999; 9905-0145086.
 PR 21-JUL-1999; 9905-0145088.
 PR 22-JUL-1999; 9905-0145085.
 PR 22-JUL-1999; 9905-0145087.
 PR 22-JUL-1999; 9905-0145089.
 PR 22-JUL-1999; 9905-0145192.
 PR 23-JUL-1999; 9905-0145145.
 PR 23-JUL-1999; 9905-0145218.
 PR 23-JUL-1999; 9905-0145224.
 PR 26-JUL-1999; 9905-0145276.
 PR 27-JUL-1999; 9905-0145913.
 PR 27-JUL-1999; 9905-0145918.
 PR 27-JUL-1999; 9905-0145919.
 PR 28-JUL-1999; 9905-0145951.
 PR 02-AUG-1999; 9905-0146386.
 PR 02-AUG-1999; 9905-0146388.
 PR 02-AUG-1999; 9905-0146389.
 PR 03-AUG-1999; 9905-0147038.
 PR 04-AUG-1999; 9905-0147204.
 PR 04-AUG-1999; 9905-0147302.
 PR 05-AUG-1999; 9905-0147192.
 PR 05-AUG-1999; 9905-0147260.
 PR 06-AUG-1999; 9905-0147303.
 PR 06-AUG-1999; 9905-0147416.
 PR 09-AUG-1999; 9905-0147935.
 PR 09-AUG-1999; 9905-0147935.
 PR 10-AUG-1999; 9905-0148171.
 PR 11-AUG-1999; 9905-0148319.
 PR 12-AUG-1999; 9905-0148341.
 PR 13-AUG-1999; 9905-0148565.
 PR 13-AUG-1999; 9905-0148684.

PR 16-AUG-1999; 9905-0149368.
 PR 17-AUG-1999; 9905-0149175.
 PR 18-AUG-1999; 9905-0149426.
 PR 20-AUG-1999; 9905-0149722.
 PR 20-AUG-1999; 9905-0149723.
 PR 20-AUG-1999; 9905-0149929.
 PR 23-AUG-1999; 9905-0149902.
 PR 23-AUG-1999; 9905-0149930.
 PR 25-AUG-1999; 9905-0150566.
 PR 26-AUG-1999; 9905-0150884.
 PR 27-AUG-1999; 9905-0151065.
 PR 27-AUG-1999; 9905-0151066.
 PR 27-AUG-1999; 9905-0151080.
 PR 30-AUG-1999; 9905-0151303.
 PR 31-AUG-1999; 9905-0151438.
 PR 01-SEP-1999; 9905-0151930.
 PR 07-SEP-1999; 9905-0152363.
 PR 10-SEP-1999; 9905-0153070.
 PR 13-SEP-1999; 9905-0153758.
 PR 15-SEP-1999; 9905-0154018.
 PR 16-SEP-1999; 9905-0154039.
 PR 20-SEP-1999; 9905-0154779.
 PR 22-SEP-1999; 9905-0155139.
 PR 23-SEP-1999; 9905-0155486.
 PR 24-SEP-1999; 9905-0155559.
 PR 28-SEP-1999; 9905-0156458.
 PR 29-SEP-1999; 9905-0156596.
 PR 04-OCT-1999; 9905-0157117.
 PR 05-OCT-1999; 9905-0157753.
 PR 06-OCT-1999; 9905-0157865.
 PR 07-OCT-1999; 9905-0158029.
 PR 08-OCT-1999; 9905-0158232.
 PR 12-OCT-1999; 9905-0158369.
 PR 13-OCT-1999; 9905-0158923.
 PR 13-OCT-1999; 9905-0158294.
 PR 13-OCT-1999; 9905-0159295.
 PR 14-OCT-1999; 9905-0159329.
 PR 14-OCT-1999; 9905-0159330.
 PR 14-OCT-1999; 9905-0159331.
 PR 14-OCT-1999; 9905-0159637.
 PR 14-OCT-1999; 9905-0159638.
 PR 18-OCT-1999; 9905-0159584.
 PR 21-OCT-1999; 9905-0160741.
 PR 21-OCT-1999; 9905-0160767.
 PR 21-OCT-1999; 9905-0160768.
 PR 21-OCT-1999; 9905-0160770.
 PR 21-OCT-1999; 9905-0160814.
 PR 21-OCT-1999; 9905-0160815.
 PR 22-OCT-1999; 9905-0160980.
 PR 22-OCT-1999; 9905-0160981.
 PR 22-OCT-1999; 9905-0160989.
 PR 25-OCT-1999; 9905-0161404.
 PR 25-OCT-1999; 9905-0161405.
 PR 25-OCT-1999; 9905-0161406.
 PR 26-OCT-1999; 9905-0161359.
 PR 26-OCT-1999; 9905-0161360.
 PR 26-OCT-1999; 9905-0161361.
 PR 28-OCT-1999; 9905-0161920.
 PR 28-OCT-1999; 9905-0161920.
 PR 28-OCT-1999; 9905-0161993.
 PR 29-OCT-1999; 9905-0162142.

Query Match 3.3%; Score 7; DB 21; Length 181;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAIFQ 191
 |||||
 Db 83 VLTAIFQ 89

RESULT 24
 AAG55907

Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAFQ 191
Db 115 VLTAFQ 121

RESULT 22

ABP26815
ID ABP26815 standard; Protein; 178 AA.

AC ABP26815;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 2806.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KV antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus pyogenes.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tetelín H;

XX WPI: 2002-352536/38.

DR N-PSDB; ABN67446.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX Claim 1; Page 3431; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

CC Sequence 178 AA;

XX Query Match 3.3%; Score 7; DB 23; Length 178;

XX Best Local Similarity 100.0%; Pred. No. 86;

XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 FGVLSDV 185
Db 123 FGVLSDV 129

RESULT 23

AAG59755
ID AAG59755 standard; Protein; 181 AA.

AC AAG59755;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 77324.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KV termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 990US-0121825.

PR 05-MAR-1999; 990US-0123180.

PR 09-MAR-1999; 990US-0123548.

PR 23-MAR-1999; 990US-0125788.

PR 29-MAR-1999; 990US-0126264.

PR 01-APR-1999; 990US-0126785.

PR 06-APR-1999; 990US-0127462.

PR 08-APR-1999; 990US-0128234.

PR 16-APR-1999; 990US-0128714.

PR 19-APR-1999; 990US-0130077.

PR 21-APR-1999; 990US-0130449.

PR 23-APR-1999; 990US-0130510.

PR 28-APR-1999; 990US-0130891.

PR 30-APR-1999; 990US-0131449.

PR 30-APR-1999; 990US-0132048.

PR 04-MAY-1999; 990US-0132484.

PR 05-MAY-1999; 990US-0132485.

PR 06-MAY-1999; 990US-0132486.

PR 07-MAY-1999; 990US-0132487.

PR 11-MAY-1999; 990US-0134256.

PR 14-MAY-1999; 990US-0134218.

PR 14-MAY-1999; 990US-0134219.

PR 14-MAY-1999; 990US-0134221.

PR 18-MAY-1999; 990US-0134370.

PR 18-MAY-1999; 990US-0134768.

PR 19-MAY-1999; 990US-0134941.

PR 20-MAY-1999; 990US-0135124.

PR 21-MAY-1999; 990US-0135353.

PR 24-MAY-1999; 990US-0135629.

PR 25-MAY-1999; 990US-0136021.

PR 27-MAY-1999; 990US-0136392.

PR 28-MAY-1999; 990US-0136782.

PR 01-JUN-1999; 990US-0137222.

PR 03-JUN-1999; 990US-0137528.

PR 04-JUN-1999; 990US-0137702.

PR 07-JUN-1999; 990US-0137724.

PR 08-JUN-1999; 990US-0138094.

PR 10-JUN-1999; 990US-0138540.

PR 10-JUN-1999; 990US-0138847.

PR 14-JUN-1999; 990US-0139119.

PR 16-JUN-1999; 990US-0139452.

PR 16-JUN-1999; 990US-0139453.

PR 17-JUN-1999; 990US-0139492.

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140961.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

3.3%; Score 7; DB 21; Length 155;

CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX Sequence 139 AA;

Query Match 3.3%; Score 7; DB 22; Length 139;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAWSADP 19

DB 59 VAWSADP 65

RESULT 20

AAU30909

ID AAU30909 standard; Protein; 152 AA.

AC AAU30909;

DT 18-DEC-2001 (first entry)

DE Novel human secreted protein #1400.

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.

PN WO200179449-A2.

PD 25-OCT-2001.

PF 16-APR-2001; 2001WO-US08656.

PR 18-APR-2000; 2000US-0552929.

PR 26-JAN-2001; 2001US-0770160.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

PI WPI; 2001-611725/70.

PS Claim 20; Page 369-370; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;

CC Immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX Sequence 152 AA;

Query Match 3.3%; Score 7; DB 22; Length 152;

Best Local Similarity 100.0%; Pred. No. 74;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAWSADP 19

DB 27 VAWSADP 33

RESULT 21

ID AAG61764 standard; Protein; 155 AA.

AC AAG61764;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 80209.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EP103405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PA (GENO-) INST GENOMIC RES.
 XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 XX WPI; 2002-352536/38.
 DR N-PSDB; ABN68279.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 PS
 PS Claim 1; Page 3608; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GBS (Streptococcus agalactiae) or group A Streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN65044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used to obtain affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 SQ Sequence 87 AA;
 XX
 OY 35 DATAITE 41
 DB 72 DATAITE 78
 XX
 RESULT 18
 AAG01133
 ID AAG01133 standard; Protein; 133 AA.
 XX
 AC AAG01133;
 XX
 DT 06-OCT-2000 (First entry)
 XX
 DE Human secreted protein, SEQ ID NO: 5214.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI; 2000-500381/45.
 DR N-PSDB; AAC01139.
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 13; SEQ ID 5214; 71pp + CD-ROM; English.
 PS
 PS The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 CC
 SQ Sequence 133 AA;
 XX
 OY 13 VAVSADP 19
 DB 59 VAVSADP 65
 XX
 RESULT 19
 AAU30907
 ID AAU30907 standard; Protein; 139 AA.
 XX
 AC AAU30907;
 XX
 DT 18-DEC-2001 (First entry)
 XX
 DE Novel human secreted protein #1398.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Dimaane RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 PS
 PS Claim 20; Page 368; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent

DR N-PSDB; AAC95476.
 XX
 XX New nucleic acid molecules encoding 50 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 XX Claim 11; Page 374; 402pp; English.
 PS
 XX polynucleotide sequences AAC95462 - AAC95511 represent cDNA encoding 50
 CC human secreted proteins AAB51929 - AAB51978. Sequences AAB51979 -
 CC AAB52010 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences with which they share homology. The genes and
 CC proteins have activities dependent on the tissues and cells in which they
 CC are expressed. Examples of their activities include immunosuppressive;
 CC antitumor; antipneumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
 CC virucide; fungicide; opthalalmological; and vulnary. The secreted
 CC proteins, polypeptides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors, and other nutritional
 CC components. Oligonucleotides AAC95453 - AAC95461 and peptide AAB51928 are
 CC used in the isolation and characterisation of the proteins and
 CC polynucleotides of the invention.
 XX
 XX Sequence 73 AA;
 SQ
 Query Match 3.3%; Score 7; DB 21; Length 73;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 155 TSFEVRQ 161
 |||||
 Db 43 TSFEVRQ 49
 RESULT 16
 AAU37687
 ID AAU37687 standard; Protein; 87 AA.
 AC
 XX AAU37687;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 XX Streptococcus pneumoniae cellular proliferation protein #116.
 DE
 XX
 XX Antisense: prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 21-MAR-2001; 2001WO-US09180.
 PF
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PA

PR 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr CJ;
 PI Yamamoto RT, Xu HH;
 PI
 XX WPI; 2001-611495/70.
 DR
 DR N-PSDB; AAS55546.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX Example 3; Seq ID No 13280; 511pp; English.
 PS
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 87 AA;
 SQ
 Query Match 3.3%; Score 7; DB 22; Length 87;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 34 DDAAIAAI 40
 |||||
 Db 71 DDAAIAAI 77
 RESULT 17
 ABP27648
 ID ABP27648 standard; Protein; 87 AA.
 AC
 XX ABP27648;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 XX Streptococcus polypeptide SEQ ID NO 4472.
 DE
 XX
 XX Streptococcus; GAS; GBS: group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-GB04789.
 PF
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 PR
 PA (CHIR-) CHIRON SPA.

```

ID AAR05898 standard; protein; 1784 AA.
XX
AC AAR05898;
XX
XX 29-NOV-1990 (first entry)
DT
XX Gene product of first ORF of virus capable of inducing viraemia in
DE feline species encoded by clone 61E.
XX
XX Viraemia; Leukaemia; Felv-A; clone 61C; clone 61E; ds.
XX
OS Feline leukaemia virus-A.
XX
PN EP377842-A.
XX
PD 18-JUL-1990.
XX
PF 12-DEC-1989; 89EP-0122964.
XX
PR 13-DEC-1988; 88US-0284139.
XX
PA (HARD ) HARVARD UNIV.
XX (COLS ) COLORADO STATE UNIV.
XX
PI Hoover EA, Mullins JI;
XX
XX MPI: 1990-218326/29.
DR N-PSDB; AAO05252.
XX
XX DNA encoding virus-A sub-type - producing AIDS type disease in
PT cats, used to test drugs and vaccines.
XX
XX Claim 6; Fig 1.1-1.7: 23pp; English.
XX
XX Variant of viral genome induces immunodeficiency in cats similar to
CC AIDS. May be used in research, especially in testing drugs and
CC vaccines against viraemia and feline leukaemia viruses.
XX
SQ Sequence 1784 AA;

Query Match 3.8%; Score 8; DB 11; Length 1784;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203
DB 513 KEMTKVLA 520

RESULT 14
AAR94427
ID AAR94427 standard; Protein; 1784 AA.
XX
AC AAR94427;
XX
XX 11-JUN-1996 (first entry)
DT
XX Felv F6a provirus clone 61E encoded nucleocapsid protein.
DE
XX Felv; retrovirus; vaccine; AIDS; disease model; immunodeficiency;
KW viraemia; leukaemia; therapy; nucleocapsid.
XX
OS Feline immunodeficiency virus subtype A.
XX
XX Key Location/Qualifiers
FH Misc-difference 581 /note="codon 581 in encoding sequence is TAG"
FT
XX
XX EP699758-A1.
PN
XX 06-MAR-1996.
PD
XX 12-DEC-1989; 89EP-0100871.
PF

```

```

XX
XX 13-DEC-1988; 88US-0284139.
PR
XX
XX (COLS ) UNIV COLORADO STATE RES FOUND.
PA (HARD ) UNIV HARVARD.
XX
XX Hoover EA, Mullins JI;
XX
XX MPI: 1996-180826/19.
DR N-PSDB; AAT13265.
XX
XX Inactivated Felv-A sub-type isolates for use in disease models and
PT vaccines - can be used to study prophylaxis and therapy of related
PR immuno-deficiencies in other species, e.g. human.
XX
XX Disclosure; Fig 1: 22pp; English.
XX
XX The nucleocapsid protein (AAR94427) and envelope protein (AAR94428)
CC sequences of feline leukaemia virus provirus F6a clone 61E were
CC deduced from open reading frames identified in the proviral DNA
CC (AAT13265). Clone 61E is replication competent and capable of
CC inducing persistent viraemia in cats. Host cells, e.g. feline cells,
CC can be transformed with DNA derived from 61E and used to produce
CC infectious virus useful in vaccines, in the generation of viraemia
CC and in disease challenge systems.
XX
SQ Sequence 1784 AA;

Query Match 3.8%; Score 8; DB 17; Length 1784;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203
DB 513 KEMTKVLA 520

RESULT 15
AAB51943
ID AAB51943 standard; Protein; 73 AA.
XX
AC AAB51943;
XX
XX 21-FEB-2001 (first entry)
DT
XX Human secreted protein sequence encoded by gene 15 SEQ ID NO:75.
DE
XX Human; secreted protein; immunosuppressive; antiarthritic; vulnery;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; hyperproliferative disorder; neoplasia;
KW autoimmune disease; rheumatoid arthritis; cardiovascular disorder;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis; viral;
KW nervous system disorder; Alzheimer's disease; bacteria; infection;
KW cell proliferation; skin aging; wound healing; chemotaxis;
KW food additive.
XX
XX Homo sapiens.
OS
XX WO2000058334-A1.
PN
XX 05-OCT-2000.
PD
XX 22-MAR-2000; 2000MO-US07507.
PF
XX 26-MAR-1999; 99US-0126594.
PR 17-DEC-1999; 99US-0172408.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
PI
XX MPI; 2000-611701/58.
DR

```

CC The invention relates to a multivalent vaccine for protecting cats
CC against several pathogens, especially pathogens associated with
CC respiratory and digestive diseases. The pathogens are especially
CC selected from feline leukaemia virus (FeLV), feline pan-leukaemia
CC virus (FPV), feline calicivirus (FCV), feline immunodeficiency virus
CC (FIV), or rabies virus. The vaccines are preferably composed of
CC polynucleotide sequences encoding 3 antigens, all as part of vectors.
CC This sequence represents the 989 protein from the feline leukaemia
CC virus (FeLV) subtype A strain Glasgow-1. The coding sequence was
CC sub-cloned into the plasmid pVR1012 to generate plasmid pPB101 for
CC use in the vaccine.
XX
SQ Sequence 580 AA;

Query Match 3.8%; Score 8; DB 19; Length 580;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KEMTKVLA 203
XX
DB 513 KEMTKVLA 520

RESULT 11
AAM50122
ID AAM50122 standard; Protein; 611 AA.
XX
AC AAM50122;
XX
DT 21-DEC-2001 (first entry)
XX
DE Feline leukemia virus p27-gp70 recombinant antigen pFeLVp27-gp70611.
XX
XX
XX FeLV; glycoprotein; gp70; p27; pFeLVp27-gp70611; antigen;
KM Immune status; vaccination status; cat.
XX
OS Feline herpesvirus.
OS Synthetic.
XX
PN WO200165568-A2.
XX
PD 13-SEP-2001.
XX
PF 07-MAR-2001; 2001WO-US07251.
XX
XX
PR 09-MAR-2000; 2000US-0521738.
XX
XX
PA (HESK-) HESKA CORP.
PA (COLS) UNIV COLORADO STATE RES FOUND.
XX
PI Jensen WA, Laplin MR, Rosen DK, Andrews JS;
XX
XX WPI; 2001-639000/73.
DR N-PSDB; AAH27069.
XX
XX
PT Determining immune status or vaccination status of an animal to e.g.
PT calicivirus comprises using a recombinant viral antigen -
PS
PS Claim 4; Page 119-121; 132pp; English.
XX
XX
XX The present sequence is that of a fusion protein, termed
CC pFeLVp27-gp70611, between the C-terminus of feline leukemia virus
CC (FeLV) pF65-gag and FeLV glycoprotein 70. The fusion protein was
CC produced in Escherichia coli cells transformed by a recombinant
CC vector comprising nucleic acid nFeLVp27-gp701833 (see AAH27069).
CC pFeLVp27-gp70611 is an example of a recombinant infectious agent
CC antigen that can be used in the method of the invention to
CC determine the immune status of an animal. The method involves
CC contacting a biological specimen of an animal (cat, dog or horse)
CC with a recombinant antigen, and detecting the presence or absence
CC of a complex between the recombinant antigen and an antibody
CC present in the sample. The method determines whether the animal is
CC protected against disease or should be vaccinated. Recombinant

CC antigens (see AAM50107-24), nucleic acids encoding them (see
CC AAH27054-71), methods of producing them, and assay methods are
CC provided.
XX
SQ Sequence 611 AA;

Query Match 3.8%; Score 8; DB 22; Length 611;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KEMTKVLA 203
XX
DB 131 KEMTKVLA 138

RESULT 12
ABB92052
ID ABB92052 standard; Protein; 783 AA.
XX
AC ABB92052;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 1263.
XX
XX
XX Herbicidal; plant; agriculture; herbicide.
KM
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
XX
XX 28-AUG-2001; 2001WO-EP09892.
PF
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX
XX (FARB) BAYER AG.
PA
PI Tietjen K, Weidner M;
PI
XX WPI; 2002-269010/31.
DR
XX
XX
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
XX
XX Claim 5; SEQ ID NO 1263; 261pp + Sequence Listing; English.
XX
XX
XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
XX
SQ Sequence 783 AA;

Query Match 3.8%; Score 8; DB 23; Length 783;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 KATIDDAIA 38
XX
DB 257 KATIDDAIA 264

RESULT 13
AAR05898

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.8%; Score 8; DB 21; Length 349;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 136 QDFVVALS 143
DB 332 QDFVVALS 339

RESULT 9
AAR41823

ID AAR41823 standard; Protein; 412 AA.

XX AAR41823;

DT 23-MAR-1994 (first entry)

DE Methylolobacillus glycogenes 1006 homoserine dehydrogenase.

KW L-threonine biosynthesis; homoserine dehydrogenase;

KW threonine synthase; homoserine kinase; amino acid; ATCC 21371.

OS Methylolobacillus glycogenes strain 1006 (ATCC 21371).

XX JP05207886-A.
XX

PD 20-AUG-1993.
XX
PF 29-JAN-1992; 92JP-0014335.
XX
PR 29-JAN-1992; 92JP-0014335.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR WPI; 1993-297465/38.
N-PSDB; AAQ48930.
XX
PT Prepn. of L-threonine by fermentation in methanol soln. - using
PT Methylolobacillus sp. contg. recombinant vector to produce
PT homoserine-dehydrogenase or kinase, or threonine synthase
XX
PS Claim 5; Page 14-15; 21pp; Japanese.
XX
CC DNA coding for at least one of the enzymes homoserine dehydrogenase,
CC threonine synthase and homoserine kinase is incorporated into a
CC recombinant expression vector and used to transform Methylolobacillus
CC bacteria. The transformants are cultured in medium containing
CC methanol as carbon source to synthesise L-threonine.
CC See also AAQ48929 and AAQ48931-Q48932.
XX
SQ Sequence 412 AA;

Query Match 3.8%; Score 8; DB 14; Length 412;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 DDAIAAIE 41
DB 386 DDAIAAIE 393

RESULT 10
AAM68402

ID AAM68402 standard; Protein; 580 AA.

XX AAM68402;

DT 28-OCT-1998 (first entry)

DE FelV-A gag protein.

KW Multivalent vaccine; cat. pathogen; respiratory disease; FelV; FCV;

KW digestive disease; feline leukaemia virus; feline pan-leukaemia virus;

KW feline calicivirus; feline immunodeficiency virus; FIV; rabies virus;

OS Feline leukemia virus.

XX FR2751223-A1.

PD 23-JAN-1998.

PF 19-JUL-1996; 96FR-0009337.

PR 19-JUL-1996; 96FR-0009337.

PA (INMR) RHONE MERIEUX SA.

PI Audonnet JCF, Baudu P, Bouchardon A, Riviere MEA;

DR WPI; 1998-112823/11.

DR N-PSDB; AAV49262.

PT Multi-valent polynucleotide vaccines against feline pathogens
PT consist of at least 3 plasmids able to express protective antigens
PT from specified viruses

XX Example 9; Fig 5; 42pp; French.
XX

PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142055.
 PR 08-JUL-1999; 99US-0142390.
 PR 09-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142820.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149829.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151348.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 3.8%; Score 8; DB 21; Length 342;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 QDFVALS 143
 |||||||
 Db 325 QDFVALS 332

RESULT 8
 AAG31330
 ID AAG31330 standard; Protein; 349 AA.
 XX
 AC AAG31330;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 37608.
 XX

```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157175.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 3.8%; Score 8; DB 21; Length 321;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 QDEVALS 143
Db 304 QDEVALS 311

RESULT 7
AAG31331
ID AAG31331 standard: Protein; 342 AA.

```
XX AC AAG31331;
XX XX 17-OCT-2000 (first entry)
DT DT
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37609.
XX XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX XX
PD PD 06-SEP-2000.
XX XX
XX XX 25-FEB-2000; 200EP-0301439.
XX XX
XX XX 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131443.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137724.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138440.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
```

DB 71 DDAIAAE 78
|||||||
RESULT 6
AAG31332
ID AAG31332 standard; Protein: 321 AA.
XX
AC AAG31332;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37610.
XX
KW Protein identification: signal transduction pathway: metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136192.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145324.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

CC allergen.
XX
SQ Sequence 215 AA;

Query Match 11.7%; Score 25; DB 15; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.1e-16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 TSFEVROFANVNHIGSLIDPIF 179
DB 155 TSFEVROFANVNHIGSLIDPIF 179

RESULT 4
AAV25586

ID AAV25586 standard; protein; 215 AA.

AC AAV25586;

DT 30-SEP-1999 (first entry)

DE D. pteronyssinus allergen Der p 7 protein fragment.

KW Major histocompatibility complex; class II; desensitizing; human;
KW allergen; grass; tree; weed; pollen; fungi; food; insect; sting;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

OS Dermatophagoides pteronyssinus.

PN WO9334826-A1.

PD 15-JUL-1999.

PF 11-JAN-1999; 99WO-GB00080.

PR 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Kay AB, Larche M;

PI WPI; 1999-458255/38.

PT Desensitizing patients to polypeptide allergens
XX
XX Example 6; Page 51; 117pp; English.

CC This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitizing patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents the house dust mite (Dermatophagoides pteronyssinus) allergen
CC Der p 7.

XX Sequence 215 AA;

SO Query Match 11.7%; Score 25; DB 20; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.1e-16;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 TSFEVROFANVNHIGSLIDPIF 179
DB 155 TSFEVROFANVNHIGSLIDPIF 179

RESULT 5
ABP27647

ID ABP27647 standard; protein; 87 AA.

AC ABP27647;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 4470.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelin H;

PI WPI; 2002-352536/38.

PI N-PSDB; ABN68278.

PS Claim 1; Page 3608; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (1), nucleic acids encoding (1), ABN6004-ABN71526 and
CC antibodies that bind (1) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (1) are used to detect Streptococcus in a
CC biological sample. (1) is used to determine whether a compound binds to
CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (1) may be used to recombinantly produce (1) and may be
CC used in gene therapy. Antibodies to (1) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

XX Sequence 87 AA;

SO Query Match 3.8%; Score 8; DB 23; Length 87;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX Claim 5; Page 40-41; 67pp; English.
 XX derYII antigen is useful as antiallergic reagent for treating
 CC sensitivity to house dust mite allergens.
 XX

SQ Sequence 213 AA;

Query Match 100.0%; Score 213; DB 15; Length 213;
 Best Local Similarity 100.0%; Pred. No. 1.9e-204;
 Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMKFLIAAFAVAVASADPIHYDKITEEINKAIDATAIEQSEETIDPMKVPDHADKFER 60
 |||||||
 DB 1 MMKFLIAAFAVAVASADPIHYDKITEEINKAIDATAIEQSEETIDPMKVPDHADKFER 60
 QY 61 HVGIVDFKGLAMRNTEARGLKOMKROGDANVGEESIVAHLLIGVHDDIVSMEDLAY 120
 |||||||
 DB 61 HVGIVDFKGLAMRNTEARGLKOMKROGDANVGEESIVAHLLIGVHDDIVSMEDLAY 120
 QY 121 KLGLDLPHTTHVIDIDIOFVALSLEISDEGNITMTSEFVQFANVYVNHIGLSILDPFIG 180
 |||||||
 DB 121 KLGLDLPHTTHVIDIDIOFVALSLEISDEGNITMTSEFVQFANVYVNHIGLSILDPFIG 180
 QY 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213
 |||||||
 DB 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213

RESULT 2
 AAY25592
 ID AAY25592 standard; protein; 213 AA.
 XX

AC AAY25592;

XX 30-SEP-1999 (first entry)

DE D. farinae allergen Der f 7 protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

XX Dermatophagoides farinae.

XX WO9934826-A1.

PD 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

DR WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 52; 117pp; English.

CC This invention describes a novel method of desensitizing a patient to a
 CC polypeptide allergen and comprises administering to the patient a peptide
 CC derived from the allergen where restriction to a MHC Class II molecule
 CC possessed by the patient can be demonstrated for the peptide and the
 CC peptide is able to induce a late phase response in an individual who
 CC possesses the MHC Class II molecule. The methods can be used for

CC desensitizing patients to allergens present in e.g. grass, tree and weed
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
 CC non-biting midge larvae, bee moth larvae, mammals such as cat, dog, horse, cow, pig,
 CC Tenidrio mollitor beetle, guinea pig, mice or gerbil. They can also be used to
 CC produce immunological vaccines which may be used to prevent and/or treat
 CC conditions involving hypersensitivity to allergens. This sequence
 CC represents the Dermatophagoides farinae allergen Der f 7.

SQ Sequence 213 AA;

Query Match 100.0%; Score 213; DB 20; Length 213;
 Best Local Similarity 100.0%; Pred. No. 1.9e-204;
 Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMKFLIAAFAVAVASADPIHYDKITEEINKAIDATAIEQSEETIDPMKVPDHADKFER 60
 |||||||
 DB 1 MMKFLIAAFAVAVASADPIHYDKITEEINKAIDATAIEQSEETIDPMKVPDHADKFER 60
 QY 61 HVGIVDFKGLAMRNTEARGLKOMKROGDANVGEESIVAHLLIGVHDDIVSMEDLAY 120
 |||||||
 DB 61 HVGIVDFKGLAMRNTEARGLKOMKROGDANVGEESIVAHLLIGVHDDIVSMEDLAY 120
 QY 121 KLGLDLPHTTHVIDIDIOFVALSLEISDEGNITMTSEFVQFANVYVNHIGLSILDPFIG 180
 |||||||
 DB 121 KLGLDLPHTTHVIDIDIOFVALSLEISDEGNITMTSEFVQFANVYVNHIGLSILDPFIG 180
 QY 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213
 |||||||
 DB 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213

RESULT 3
 AAR60575
 ID AAR60575 standard; protein; 215 AA.
 XX

AC AAR60575;

XX 01-APR-1995 (first entry)

DE House dust mite allergen DerpVII cDNA.

XX derpVII allergen; antiallergic; allergy diagnosis.

XX Dermatophagoides pteronyssinus.

XX WO9420614-A.

PD 15-SEP-1994.

XX 11-MAR-1994; 94WO-AU00117.

XX 12-MAR-1993; 93US-0031141.

PR 22-JUN-1993; 93US-0081540.

XX (CHIL-) INST CHILD HEALTH RES.

XX Chua K, Thomas WR;

DR WPI; 1994-303021/37.

DR N-PSDB; AAO71400.

XX New nucleic acid encoding specific dust mite allergens - and
 XX related vectors, transformed cells, peptides and antibodies,
 XX useful for desensitisation and diagnosis.

PS Claim 7; Page 36-37; 67pp; English.

CC DerpVII antigen is useful as antiallergic reagent for treating
 CC sensitivity to house dust mite allergens. The DNA can be used
 CC as a probe to detect the sensitivity of an individual to the

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:09:45 ; Search time 35 Seconds
(without alignments)
810.925 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213
Sequence: 1 MMKFLIAAVAFVAVSADPI.....VRKENTKYLAPAFKRELEKN 213

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	213	15 AAR60576	House dust mite al
2	213	100.0	213	20 AAY25592	D. farinae allerg
3	213	11.7	215	15 AAR60575	House dust mite al
4	25	11.7	215	20 AAY25586	D. pteronyssinus a
5	8	3.8	87	23 ABP27647	Streptococcus poly
6	8	3.8	321	21 AAG31332	Arabidopsis thalia
7	8	3.8	342	21 AAG31331	Arabidopsis thalia
8	8	3.8	349	21 AAG31330	Arabidopsis thalia
9	8	3.8	412	14 AAR41823	Methylobacillus gl
10	8	3.8	580	19 AAW68402	Felvy-A gag protein

11	8	3.8	611	22 AAM50122	Feline leukemia vi
12	8	3.8	783	23 ABB92052	Herbicidally activ
13	8	3.8	1784	11 AAR05898	Gene product of fi
14	8	3.8	1784	17 AAR94427	Felvy F6a provirus
15	7	3.3	73	21 AAB51943	Human secreted pro
16	7	3.3	87	22 AAB27648	Streptococcus pneu
17	7	3.3	133	23 AAG01133	Human secreted pro
18	7	3.3	139	22 AAV30907	Novel human secret
19	7	3.3	152	22 AAV30909	Novel human secret
20	7	3.3	155	21 AAG61764	Arabidopsis thalia
21	7	3.3	178	23 AAB26815	Streptococcus poly
22	7	3.3	181	21 AAG59755	Arabidopsis thalia
23	7	3.3	183	21 AAG55907	Arabidopsis thalia
24	7	3.3	183	21 AAG59754	Arabidopsis thalia
25	7	3.3	183	21 AAG59754	Arabidopsis thalia
26	7	3.3	185	21 AAG55324	Arabidopsis thalia
27	7	3.3	193	23 AAB06132	Human NS protein s
28	7	3.3	204	15 AAR54114	GDI D4 protein. B
29	7	3.3	205	21 AAG59753	Arabidopsis thalia
30	7	3.3	214	21 AAG55323	Arabidopsis thalia
31	7	3.3	215	21 AAG55906	Arabidopsis thalia
32	7	3.3	300	21 AAB18216	Plasmodium falci
33	7	3.3	354	14 AAR33439	Ornithine cyclodea
34	7	3.3	357	14 AAR33440	Lactococcus lactis
35	7	3.3	434	23 ABB54464	Drosophila melanog
36	7	3.3	460	22 ABB63535	Herbicidally activ
37	7	3.3	474	23 ABB92783	Drosophila melanog
38	7	3.3	511	22 ABB62629	Drosophila melanog
39	7	3.3	581	22 ABB25418	Novel human diagno
40	7	3.3	752	14 AAR38153	Acetobacter di
41	7	3.3	791	23 ABB93713	Herbicidally activ
42	7	3.3	2836	22 ABB62719	Drosophila melanog
43	6	2.8	8	22 AAB23790	Voltage sensitive
44	6	2.8	10	16 AAR70305	Laminin (LM) bindi
45	6	2.8	10	23 ABB67586	Human ADP1 tryptic

ALIGNMENTS

RESULT 1
ID AAR60576 standard; Protein: 213 AA.
AC AAR60576;
DT 01-APR-1995 (first entry)
XX House dust mite allergen DerfVII.
DE DerfVII allergen; anti-allergic; allergy diagnosis.
XX Dermatophagoides farinae.
OS WO9420614-A.
XX
PN 15-SEP-1994.
PD
XX 11-MAR-1994; 94WO-AU00117.
PF
XX 12-MAR-1993; 93US-0031141.
PR 22-JUN-1993; 93US-0081540.
XX
PA (CHIL-) INST CHILD HEALTH RES.
XX Chua K, Thomas WR;
XX WPI: 1994-303021/37.
XX N-PSDB; AAQ71401.
XX New nucleic acid encoding specific dust mite allergens - and
XX related vectors, transformed cells, peptides and antibodies,
XX useful for desensitisation and diagnosis.
PT

Fri Feb 21 10:04:03 2003

us-10-024-955-7_1.rst

Page 33

Oy 103 LeuLeuIleGlyValAlHisAspAPLevalSermeCclUtyrAspleuAtaryTlysLeu 122
 ||| ||| ||| ||| ||| |||
Db 400 GCACTGGAAAGGTACTACAGGCACAGACTCTTTGACCAGCGGGAGCTCCCTTTGAGATT 459

Oy 123 Glyasp 124
 |||
Db 460 GGCGAG 465

Search completed: February 20, 2003, 17:29:40
Job time : 1492 secs

Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.29% Indels: 17
 DB: 14 Gaps: 3

US-10-024-955-7 (1-213) x B0689372 (1-919)

```

OY 18 AspproIleHstYrAspLysIleThrgIuIleAsnLysAlaIleAspAlaIle 37
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 333 GACAAATTCTCTTACCTGGGAAAGGAGGATCATGACGCTTGACATTGCCATA 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 38 AlaAlaIleGluGlnSerGluThrIle----- 46
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 393 GCCACCATGAAAGGTGGGAGGTGTGCCACATCACCCTGCAACAGAAATATGCTACGGT 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 47 -----AspprometLysValProAspHisAlaAsp---LysPheGluArgHisVal 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 453 TCAGCAGCAGCAGTCCTCCAAAGATTCGCCCAATGCCACCTGTATTTAG-----GTG 506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 63 GlyIleValAspPheLysGlyLysIleValMetArgAsnIleGluAlaArgGlyLeuLys 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 507 GAGTTGTTGAGTTTAAGGAGAGATCTGCAGGAAGAGAGATGCCGAATCATTCGCC 566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLysAlaHis 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 567 AGAATACAGACTCGCGTGAAGCTATGCTAAGCCCAATGAGGCTGTATGCGAGATT 626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuLaryLysLeu 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 627 GCACGTGGAAGGTACTACAAAGCAAGCTTTTGACCACGGAGGCTCGCTTGAGATT 686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 123 GlyAsp 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 687 GCGCAG 692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 46
 BGI06670 920 bp mRNA linear EST 30-JAN-2001
 LOCUS 602290509F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4385187 5',
 DEFINITION mRNA sequence.
 BGI06670
 ACCESSION BGI06670.1 GI:12600516
 VERSION BGI06670.1 GI:12600516
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 920)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Louis Staudt, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM1006 row: b column: 04
 High quality sequence stop: 707.
 Location/Qualifiers
 1..920

FEATURES
 source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4385187"
/clone_11b="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Lymph. Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies."

```

BASE COUNT 239 a 231 c 265 g 185 t
 ORIGIN
 Note: this is a NIH_MGC library."

Alignment Scores:

Pred. No.:	0.494	Length:	920
Score:	88.30	Matches:	30
Percent Similarity:	45.08%	Conservative:	25
Best Local Similarity:	24.59%	Mismatches:	50
Query Match:	8.29%	Indels:	17
DB:	12	Gaps:	3

US-10-024-955-7 (1-213) x BGI06670 (1-920)

```

OY 18 AspproIleHstYrAspLysIleThrgIuIleAsnLysAlaIleAspAlaIle 37
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
DB 178 GACAAATTCTCTTACCTGGGAAAGGAGGATCATGACGCTTGACATTGCCATA 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 38 AlaAlaIleGluGlnSerGluThrIle----- 46
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 238 GCCACCATGAAAGGTGGGAGGTGTGCCACATCACCCTGCAACAGAAATATGCTACGGT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 47 -----AspprometLysValProAspHisAlaAsp---LysPheGluArgHisVal 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 298 TCAGCAGCAGCAGTCCTCCAAAGATTCGCCCAATGCCACGCTGTATTTAG-----GTG 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 63 GlyIleValAspPheLysGlyLysIleValMetArgAsnIleGluAlaArgGlyLeuLys 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 GAGTTGTTGAGTTTAAGGAGAGATCTGCAGGAAGAGAGATGCCGAATCATTCGCC 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLysAlaHis 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 412 AGAATACAGACTCGCGTGAAGCTATGCTAAGCCCAATGAGGCTGTATGCGAGATT 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuLaryLysLeu 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 472 GCACGTGGAAGGTACTACAAAGCAAGCTTTTGACCACGGAGGCTCGCTTGAGATT 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 123 GlyAsp 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 532 GCGCAG 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 47
 AL558324 921 bp mRNA linear EST 16-FEB-2001
 LOCUS AL558324 LTI_NFL008_TC2 Homo sapiens cDNA clone CS0DJ005YN18 5
 DEFINITION prime, mRNA sequence.
 AL558324
 ACCESSION AL558324
 VERSION AL558324.1 GI:12902737
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 921)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..921

FEATURES
 source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DJ005YN18"
/clone_11b="LTI_NFL008_TC2"
/sex="male"
/tissue_type="T cells from T cell leukemia"
/notes="Vector: pCMVSPORT 6; Site:1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and

```

```

Qy 63 GylleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuys 82
      ::::::::::::::::::::
Db 511 GAGTGTGAGTTAAAGGAGAGATCTGACGGAAGAGGAGCGGAGTCATTGCG 570
Qy 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyGlyLeValLysAlaHis 102
      ::::::::::::::::::::
Db 571 AGAATACAGACTCGCGGAGGCTATGCTAGCCCAATAGAGGTCGTATCGTGAGGTT 630
Qy 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
      ::::::::::::::::::::
Db 631 GCACGTGGAAGGCTACTACAGACACAGCTCTTGACACGCGGAGCTCCGCTTGAGATT 690
Qy 123 GlysAsp 124
      ::::::::::
Db 691 GGCAG 696

RESULT 44
BQ431048 912 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7828682 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6056472
DEFINITION 5', mRNA sequence.
ACCESSION BQ431048
VERSION BQ431048.1 GI:21170124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 912)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DPF/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LLM3318 row: g column: 01
High quality sequence stop: 609.
FEATURES
source
location/Qualifiers
1..912
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6056472"
/clone_id="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 256 a 195 c 262 g 199 t
ORIGIN
Alignment Scores:
Pred. No.: 0.487 Length: 912
Score: 88.50 Matches: 30
Percent Similarity: 45.08% Conservative: 25
Best Local Similarity: 24.59% Mismatches: 50
Query Match: 8.29% Indels: 17
DB: 14 Gaps: 3
US-10-024-955-7 (1-213) x BQ431048 (1-912)
Qy 18 AspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAlaIle 37
      ::::::::::::::::::::
Db 128 GACCAATTCCTTCCTTGACCTGGGAAGGAGGTCATCAAGCGCTTGACATTGCCATA 187
Qy 38 AlaAlaIleGluGlnSerGluThrIle----- 46

```

```

Db 188 GCCACCATCAAGCTGGGGAGGTGCCACATCCACCTGCAACACGAAATATGCTACGGT 247
      ::::::::::::::::::::
Qy 47 -----AspProMetLysValProAspHisAlaAsp---LysPheGluArgHisVal 62
      ::::::::::::::::::::
Db 248 TCAGCAGCAGCTCCGCCAAGATTCGCCCAATGCTCCAGCTTGATTGAG-----GTG 301
Qy 63 GylleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuys 82
      ::::::::::::::::::::
Db 302 GAGTGTGAGTTAAAGGAGAGATCTGACGGAAGAGGAGCGGAGTCATTGCG 361
Qy 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyGlyLeValLysAlaHis 102
      ::::::::::::::::::::
Db 362 AGAATACAGACTCGCGGAGGCTATGCTAGCCCAATGAGGTCGTATCGTGAGGTT 421
Qy 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
      ::::::::::::::::::::
Db 422 GCACGTGGAAGGCTACTACAGACACAGCTCTTGACACGCGGAGCTCCGCTTGAGATT 481
Qy 123 GlysAsp 124
      ::::::::::
Db 482 GGCAG 487

RESULT 45
BQ689372 919 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8341886 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6248764
DEFINITION 5', mRNA sequence.
ACCESSION BQ689372
VERSION BQ689372.1 GI:21814688
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 919)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LLM2389 row: c column: 05
High quality sequence stop: 702.
FEATURES
source
location/Qualifiers
1..919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6248764"
/clone_id="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCGACAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 230 a 227 c 290 g 171 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 0.493 Length: 919
Score: 88.50 Matches: 30
Percent Similarity: 45.08% Conservative: 25

```

REFERENCE 1 (bases 1 to 899)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLM13612 row: b column: 05
 High quality sequence stop: 698.

FEATURES

Source

1. 899
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6199204"
 /clone_lib="Lupski_sciatic_nerve"
 /sex="male"
 /tissue_type="sciatic nerve"
 /dev_stage="adult, 70 yr"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1: NotI, Site 2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCACGCGCTCG-3' and 5'-GACGTGTTCTAGATCGAGCGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.87 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."
 Location/Qualifiers

BASE COUNT 254 a 199 c 250 g 194 t 2 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.477 Length: 899
 Score: 88.50 Matches: 30
 Percent Similarity: 45.08% Conservative: 25
 Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.29% Indels: 17
 DB: 14 Gaps: 3

US-10-024-955-7 (1-213) x B0940104 (1-899)

OY 18 ASPPROIIEHISTYFASPLYSILETHRGUIGUILEASNLYSALALEASPAALAIIE 37
 ||| :||| ||||| ||||| ||| |||||
 DB 31 GACAAATTTCTCTTACCTTGGGAAAGGGGAGGTATCAAGGCTTGGACATTGCCATA 90
 OY 38 AAlaAlaIleGluGlnSerGluThrIle----- 46
 ||| :||| |||
 DB 91 GCCACCATGAAGGTGGGGAGGTGGCCACATCACCTGCACAACCAAGATATGCTACGCT 150
 OY 47 -----AsPROMETLYSValProAspHisAlaAsp---LysPheGluArgHisVal 62
 ||| :||| ||| :||| ||| :||| |||
 DB 151 TCAGCAGCGAGTCCTCCCAAGATTCGCCCAATGCCACGCTGTATTGTAG-----GTG 204
 OY 63 GAlIleValAsPheLysGluGlnGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLys 82
 :||| :||| ||||| ||||| ||| :|||
 DB 205 GAGTGTGTTGAATTTAAGGAGAAAGATCTGACGGAAGAGGAGATGCCGAATCATTCGC 264
 OY 83 GImetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyLLeuValLysAlaHis 102
 :||| :||| ||| :||| ||| :||| ||| :|||
 DB 265 AGAATACAGACTCCGGGTGAAGCTATGCTAAGCCCAATGAGGAGGCTATCGTGAAGTT 324
 OY 103 LeuIleGlyValHisAspArgIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
 ||| :||| ||| :||| ||| :||| ||| :|||
 DB 325 GCACGTGAAGGGTACTACCAAGCAAGCTCTTGCACACGGGAGCTCGCTTGAGATT 384

OY 123 GLYAsp 124
 ||| :|||
 DB 385 GGGGAG 390

RESULT 43
 B0063355
 LOCUS B0063355
 DEFINITION AGENCOURT_6876547 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5924953
 5', mRNA sequence.
 ACCESSION B0063355
 VERSION B0063355.1 GI:19891019
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 905)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Lou Staudt
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LCM2095 row: 0 column: 02
 High quality sequence stop: 643.

FEATURES

Source

1. 905
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5924953"
 /clone_lib="NIH_MGC_99"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
 Location/Qualifiers

BASE COUNT 224 a 229 c 287 g 165 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.482 Length: 905
 Score: 88.50 Matches: 30
 Percent Similarity: 45.08% Conservative: 25
 Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.29% Indels: 17
 DB: 14 Gaps: 3

US-10-024-955-7 (1-213) x B0063355 (1-905)

OY 18 ASPPROIIEHISTYFASPLYSILETHRGUIGUILEASNLYSALALEASPAALAIIE 37
 ||| :||| ||||| ||||| ||| |||||
 DB 337 GACAAATTTCTCTTACCTTGGGAAAGGGGAGGTATCAAGGCTTGGACATTGCCATA 396
 OY 38 AAlaAlaIleGluGlnSerGluThrIle----- 46
 ||| :||| |||
 DB 397 GCCACCATGAAGGTGGGGAGGTGGCCACATCACCTGCACAACCAAGATATGCTACGCT 456
 OY 47 -----AsPROMETLYSValProAspHisAlaAsp---LysPheGluArgHisVal 62
 ||| :||| ||| :||| ||| :||| ||| :|||
 DB 457 TCAGCAGCGAGTCCTCCCAAGATTCGCCCAATGCCACGCTGTATTGTAG-----GTG 510

/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 215 a 227 c 283 g 158 t
ORIGIN

Alignment Scores:

Pred. No.: 0.464 Length: 883
Score: 88.50 Matches: 30
Percent Similarity: 45.08% Conservative: 25
Best Local Similarity: 24.59% Mismatches: 50
Query Match: 8.29% Indels: 17
DB: 14 Gaps: 3

US-10-024-955-7 (1-213) x B068609 (1-883)

OY 18 AspprolleHSTYrAspLysIleThrcIugIuIleAsnLysAlaIleAspAspAlaIle 37
Db 310 GACAAATTCTCTTGACCTGGGAAAGGGAGGTCATCAAGGCTTGAGCATGCCATA 369
OY 38 AlaAlaIleGluGlnSerGluThrIle----- 46
Db 370 GCCACCAATGAAAGTGGGGAGGTGCCACATCACCCTGCAACCAATATCCCTACGTT 429
OY 47 -----AspprometLysValProAspHisAlaAsp---LyspHeGluArghIsvAl 62
Db 430 TCAGCAGGAGAGCTCCCAAGATTCGCCCATGCGCTGTATTGAG-----GTG 483
OY 63 GlyIleValAspPheLysGlyIleuAlaMetArgAsnIleGluAlaArgIleuLys 82
Db 484 GAGTTGTTGAGTTTAAGGAGAAAGATCGACGGAAGAGAAATGCGGGAATCATTCGCG 543
OY 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLysAlaHis 102
Db 544 AGAATACAGACTCGCGGTGATGCTATAGCCCAATGAGGCTGTCTGTGAGAGTT 603
OY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
Db 604 GCACGTGAAGGCTACTACAGACAAAGCTCTTGACCAAGCGGAGACTCCGCTTGAGATT 663
OY 123 GlyAsp 124
Db 664 GCGGAG 669

RESULT 41
LOCUS BE795760 888 bp mRNA linear EST 20-SEP-2000
DEFINITION 601590780F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945067 5',
mRNA sequence.
ACCESSION BE795760
VERSION BE795760.1 GI:10216958
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 888)
TITLE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: csaeb@r-mail.nih.gov
Tissue Procurement: DCM/DHP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: L1CM802 row: 0 column: 20
High quality sequence stop: 773.

FEATURES

location/Qualifiers
1..888

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3945067"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 252 a 181 c 260 g 195 t
ORIGIN

Alignment Scores:

Pred. No.: 0.468 Length: 888
Score: 88.50 Matches: 30
Percent Similarity: 45.08% Conservative: 25
Best Local Similarity: 24.59% Mismatches: 50
Query Match: 8.29% Indels: 17
DB: 12 Gaps: 3

US-10-024-955-7 (1-213) x BE795760 (1-888)

OY 18 AspprolleHSTYrAspLysIleThrcIugIuIleAsnLysAlaIleAspAspAlaIle 37
Db 77 GACAAATTCTCTTGACCTGGGAAAGGGAGGTCATCAAGGCTTGAGCATGCCATA 136
OY 38 AlaAlaIleGluGlnSerGluThrIle----- 46
Db 137 GCCACCAATGAAAGTGGGGAGGTGCCACATCACCCTGCAACCAATATCCCTACGTT 196
OY 47 -----AspprometLysValProAspHisAlaAsp---LyspHeGluArghIsvAl 62
Db 197 TCAGCAGGAGAGCTCCCAAGATTCGCCCATGCGCTGTATTGAG-----GTG 250
OY 63 GlyIleValAspPheLysGlyIleuAlaMetArgAsnIleGluAlaArgIleuLys 82
Db 251 GAGTTGTTGAGTTTAAGGAGAAAGATCGACGGAAGAGAAATGCGGGAATCATTCGCG 310
OY 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLysAlaHis 102
Db 311 AGAATACAGACTCGCGGTGATGCTATAGCCCAATGAGGCTGTCTGTGAGAGTT 370
OY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
Db 371 GCACGTGAAGGCTACTACAGACAAAGCTCTTGACCAAGCGGAGACTCCGCTTGAGATT 430
OY 123 GlyAsp 124
Db 431 GCGGAG 436

RESULT 42
LOCUS B0940104 899 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8841415 Lupsk1.scf1a1c.neive Homo sapiens cDNA clone
IMAGE:6199204 5', mRNA sequence.
ACCESSION B0940104
VERSION B0940104.1 GI:22355582
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OY	97	GlyIleValIysAlaHisLeuLeuGluValAlaHis-----AspAspIleValSerMet	114
Db	421	---TTGACCAATGGCATACTCATCCATGAGCGCTGCCACCATGATGAATGCCTTAGCGCCTT	365
OY	115	GIuTYrAspLeuAlaIleTYrLysLeuGluYAspLeuHisProThrThrHisValIleSerAsp	134
Db	364	CTTCCTTACGCTTCTCCAGAAGAGGAGAGTGTTCCACAGCGCTTCTTGCTGTCCACAGTATT	305
OY	135	IleGlnAspPheValValAlaLeuSerLeuGluTIIeSerAspGluGlyAsnIleThrMet	154
Db	304	GTAGTAAGTCTTCGCTT-----CTGACCCCTTCCTTACATCCGTGTACATAGTCCTT	257
OY	155	ThrSerPheGluValArgGlnPheAlaIasValValaHisIleGlyGly-----	171
Db	256	-----GAGCGCTCGTAC-----CTCATCACACACTCTTGCTGGAGTGTACT	215
OY	172	-----LeuSerIleLeuAspProIlePheGlyValLeuSerAsp-----	184
Db	214	CAGAAAGCTCAGCAATCTTGCTCTTCTGCGAGCTCATGTAGATGCCAAGCTTGAGTT	155
OY	185	-----ValLeuThrAlaIlePheGlnAspThrValArgIys	196
Db	154	CTTGAGAGAAGCGCTCGTAGAACCTTGTGTAGTCCTCTTGCTCCAGCAATCTCAAGAA	95
OY	197	GIUmetThrIysValLeuAlaProAlaPheIysArGluLeuGlu	211
Db	94	GAGCTCAATGCACCTTGTGACAAGGTTCTTCGGATGACTTGAG	50
RESULT 39	BQ923854	867 bp	mRNA linear EST 20-AUG-2002
LOCUS	BQ923854		
DEFINITION	AGENCOURT_0806910 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:633833		
ACCESSION	. BQ923854		
VERSION	BQ923854.1 GI:22338885		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 867)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.lnl.gov Plate: LMC2531 row: g column: 06 High quality sequence stop: 658. Location/Qualifiers 1..867 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="633833" /clone_id="NIH_MGC_47" /tissue_type="neuroblastoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: brain; Vector: pOT8; site.1: XhoI; site.2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5'- adapter: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC Library."		
FEATURES	source		
BASE COUNT	237 a 191 c 273 g 166 t		

[illegible]

US-10-024-955-7 (1-213) x Bf136676 (1-686)

```

Oy 18 AspproillehstYrAspysliethrgluileasnllyalalleaspaspaiaile 37
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 10 GACAAATTCCTCTTGGACCTGGGAAAAGGGAGATCATCAAGGCTTGGATATTCCTGTG 69
Oy 38 AlaAlailegluInsergluThrile----- 46
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 70 GCAACCATGAAGTGGGGAGGTGTGCCACATCACCTGCAGGCACATATGCTATGGC 129
Oy 47 -----AsppromellysValProasphIsAlaAsp---LysPhegluarghIsVal 62
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 130 GCAGCAGCGACGCTCCGGAAGATCCGCCCAACGCCACATCTGTATTTAG-----GTG 183
Oy 63 GlyIleValAspPheylsglyIleuAlaMetArgAsnIleGluAlaArglyLeuLys 82
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 184 GAGCTGTTGATTCACAAAGAGAACCTTACAGAAAGAAAGATGCGCGATCATCCGC 243
Oy 83 GlmetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyIleValLysAlaHis 102
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 244 ACAATATGACACTCGGGGTACAGGCTATGCCAGGCCCAATGATGCTGTATGGTGAAGTG 303
Oy 103 LeuLeuIleGlyValHisAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 304 GCCCTGGAAGCTACCAAGACGCCCTTGTGACACGCGGAGCTTGTGTAAGTC 363
Oy 123 GlyAsp 124
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 364 GGGGAA 369

```

RESULT 36
LOCUS B1917792 688 bp mRNA linear EST 16-OCT-2001
DEFINITION 603183892P1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5247815 5',
mRNA sequence.
ACCESSION B1917792
VERSION B1917792.1 GI:16181551
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT Unpublished (1999) National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L16M1625 row: h column: 24
High quality sequence start: 3
High quality sequence stop: 673.
Location/Qualifiers

FEATURES
source 1..688
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5247815"
/clone.lib="NIH_MGC_121"
/lab.host="DH10B"

/note="Organ: Brain; Vector: pCMV-SPORT6; Site_1: NOTI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."

BASE COUNT 165 a 180 c 225 g 118 t
ORIGIN

Alignment Scores:
Pred. No.: 0.315 Length: 688
Score: 88.50 Matches: 30
Percent Similarity: 45.08% Conservative: 25
Best Local Similarity: 24.59% Mismatches: 50
Query Match: 8.29% Indels: 17
DB: 13 Gaps: 3

US-10-024-955-7 (1-213) x B1917792 (1-688)

```

Oy 18 AspproillehstYrAspysliethrgluileasnllyalalleaspaspaiaile 37
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 327 GACAAATTCCTCTTGGACCTGGGAAAAGGGAGATCATCAAGGCTTGGATATTCCTGTG 386
Oy 38 AlaAlailegluInsergluThrile----- 46
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 387 GCCACCATGAAGTGGGGAGGTGTGCCACATCACCTGCAGGCACATATGCTACGGT 446
Oy 47 -----AsppromellysValProasphIsAlaAsp---LysPhegluarghIsVal 62
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 447 TCAGCAGCGACGCTCCCAAGATTCGCCCAATGACCGCTGTATTTAG-----GTG 500
Oy 63 GlyIleValAspPheylsglyIleuAlaMetArgAsnIleGluAlaArglyLeuLys 82
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 501 GAGTGTGTTGATTAAGGAGAGATCTGACCGAAGAGAAAGATGCGGAATCATTCGC 560
Oy 83 GlmetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyIleValLysAlaHis 102
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 561 ACAATATGACACTCGGGGTACAGGCTATGCCAGGCCCAATGATGCTGTATGGTGAAGTG 620
Oy 103 LeuLeuIleGlyValHisAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 621 GCACGTGAAGGCTACCAAGACAGCTTGTGACACGCGGAGCTCCGCTTGAAGTC 680
Oy 123 GlyAsp 124
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 681 GGGGAG 686

```

RESULT 37
LOCUS BG775260 707 bp mRNA linear EST 15-MAY-2001
DEFINITION 602650274P1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:4761162 5',
mRNA sequence.
ACCESSION BG775260
VERSION BG775260.1 GI:14045577
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT Unpublished (1999) National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L16M1613 row: c column: 19
High quality sequence stop: 705.
Location/Qualifiers

FEATURES
source 1..707
/organism="Homo sapiens"

Db 503 AGAATACAGACTCGCGTGAAGGCTATGCTAAGCCCAATGAGGCTGCTATCGTGAGGTT 562
 QY 103 LeuLeuIleGlyValHisaspPileValSerMetGluTyrAspLeuAlaIleLysLeu 122
 Db 563 GCACGTGAGAGGCTACTACAGACAAAGCTCTTTGACACGGGAGGCTCCGCTTTGAGATT 622
 QY 123 GlyAsp 124
 Db 623 GCGCAG 628
 RESULT 34
 LOCUS Bg748642 678 bp mRNA linear EST 15-MAY-2001
 DEFINITION 602706344P1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4843043 5',
 mRNA sequence.
 ACCESSION Bg748642
 VERSION Bg748642.1 GI:14059295
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 678)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM1678 row: g column: 12
 High quality sequence stop: 678.
 Location/Qualifiers
 1..678
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4843043"
 /clone_lib="NIH_MGC_43"
 /tissue_type="normal pigmented retinal epithelium"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pOT87; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

BASE COUNT 168 a 169 c 223 g 118 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.308 Length: 678
 Score: 88.50 Matches: 30
 Percent Similarity: 45.08% Conservative: 25
 Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.29% Indels: 17
 DB: 12 Gaps: 3
 US-10-024-955-7 (1-213) x Bg748642 (1-678)

QY 18 AspProIleHisTyrAspLysIleThrGluLeuLeuAsnLysAlaIleAspAlaIle 37
 Db 311 GACAATTTCTCTTACCTGGGAAAGGAGGAGCTATCAAGCTTGGGACATTCGCATA 370
 QY 38 AlaAlaIleGluGlnSerGluThrIle----- 46
 Db 371 GCCACCATGAGGTGGGAGAGTGTCACATCACCTCGAACCAGATATATGCTACGGT 430

QY 47 -----AspProMetLysValProAspHisAlaAsp---LysPheGluArgHisVal 62
 Db 431 TCAGCAGGAGGAGTCCTCCAAAGATTCCTCCCATTCGCACGCTGTATTGAG-----GTG 484
 QY 63 GlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgLysLeu 82
 Db 485 GAGTTGTGAGCTTTAAGCGAGAGATCTGACGGAAGAGAGAGATGCGGATTCATTCGC 544
 QY 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlnGlyIleValLysAlaHis 102
 Db 545 AGAATACAGACTCGCGTGAAGCTATGCTAAGCCCAATGAGGCTGCTATCGTGAGGTT 604
 QY 103 LeuLeuIleGlyValHisaspPileValSerMetGluTyrAspLeuAlaIleLysLeu 122
 Db 605 GCACGTGAGAGGCTACTACAGACAAAGCTCTTTGACACGGGAGGAGCTCCGCTTTGAGATT 664
 QY 123 GlyAsp 124
 Db 665 GCGCAG 670
 RESULT 35
 LOCUS Bf136676 686 bp mRNA linear EST 24-OCT-2000
 DEFINITION 601780441P1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4008466 5',
 mRNA sequence.
 ACCESSION Bf136676
 VERSION Bf136676.1 GI:10975716
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 686)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLM9243 row: a column: 11
 High quality sequence stop: 684.
 Location/Qualifiers
 1..686
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="4008466"
 /clone_lib="NCI_CGAP_Lu30"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; transgenic model MMTV-LTR, expression driven by
 MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
 dT. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 179 a 154 c 226 g 126 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.314 Length: 686
 Score: 88.50 Matches: 28
 Percent Similarity: 45.90% Conservative: 28
 Best Local Similarity: 22.95% Mismatches: 49
 Query Match: 8.29% Indels: 17
 DB: 12 Gaps: 3

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 623)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapsr@email.nih.gov Tissue Procurement: DCM/DTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LCM1931 row: 1 column: 18 High quality sequence stop: 623. Location/Qualifiers
FEATURES	1..623
SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5447585" /clone_id="NIH-MGC_40" /tissue_type="carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: prostate; Vector: pOP7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed Ling Hong in the laboratory of Gerald M. Rubin (Univers of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."
BASE COUNT	164 a 135 c 193 g 131 t
ORIGIN	
Alignment Scores:	
Pred. No.:	0 27 Length: 623
Score:	88.50 Matches: 30
Percent Similarity:	45.08% Conservative: 25
Best Local Similarity:	24.59% Mismatches: 50
Query Match:	8.29% Indels: 17
DB:	13 Gaps: 3
US-10-024-955-7 (1-213) x BK044473 (1-623)	
QY	18 ASPROILENISTYRASPPLYSILETHRGUCLUJLEASNYLSALALEASPAPALALE 37
Db	146 GACAAATCTCTCTTGACACTGGGAAAGGGAGGTCTATCAAGCTTGGACATTGCCATA 205
QY	38 AAlaAlaIleGIuInSerGIuThrIle----- 46
Db	206 GCCACACATGAAGGCTGGGGAGAGTGTGCCACATCATCTGCAAAACCGAATATGCGTTACGGT 265
QY	47 -----ASPRometIlySvaIProAspIhIlaSvP-----LysPheGIuATGhISvaI 62
Db	266 TCACAGGAGCACTCTCCCAAGATCTCCCAATGTCCACCGCTTGATTGAG-----GTG 319
QY	63 GLYIleValaSPheIySgIyGluLeuAlaMetArGAsnIleGIuAlaIaAGIyLeuLys 82
Db	320 GAGTTGTTTGAAGGGAGGAAGATCTGCACGAAGAGMAATGGCGGAATCATTCGCG 379
QY	83 GImeIySaIySgInGIyASpAlaSnValIySgIyGluGIuGIuIleValIySAlaIhIS 102
Db	380 AGAATACAGACTCCCGGTGAAGGCTATGTAAGCCCAATGAGGGGTGCTATCTGGAGGTT 439
QY	103 LeuIleuIleGIyValIhISAspAspIleValISerMetGIuIyurAspLeuAlaIyTrIyLeu 122
Db	440 GCATGTGAAGGGTACTACAGGACACACTTTTGACCAGCGGGAGCTCCGCTTGAAGATT 499
Y	123 GLyAsp 124

US-10-024-955-7 (1-213) x BM61957 (1-929)

```

OY 18 Asppro11eHstYrAspLysIleThrGluIleAsnLysAlaIleAspAspAlaIle 37
    ||| ::::: |||::: ||| ||| ||| |||
Db 379 GACAAATTCCTCTTGAACCTCGGAGAAAGGAGGATCATCAAGCTTGGACATTTGCCATA 438
OY 38 AlaAla11eGluInserGIuThrIle----- 46
    ||| ::::: |||
Db 439 GCCACATGAAGTGGGAGGTGTGCCACATCACTGCAAAACAGATATGCTACGCT 498
OY 47 -----AspPrometLysValProAspHisAlaAsp---LysPheGluArgHisVal 62
    ||| |||::: |||::: ||| ||| ||| |||
Db 499 TCAGCAGCAGCAGCTCCCAAGATTTCCCAATGCCACAGCTGTATTTGAG-----GTG 552
OY 63 GlyIleValAspPheLysGlyGluLeuAlaMetArgAsn11eGluIleArgGlyLeuLys 82
    ::::: ||| ||| ||| ||| ||| |||
Db 553 GACTGTGTTGATTTAAAGGAGAGATCTGACGAGAGAGAGATGGCGGATTCATTCGCG 612
OY 83 GlmetLysArgGlnGlyAspAlaAsnValLysGlyGluGluIleValLysAlaHis 102
    ::::: ||| ||| ||| ||| ||| |||
Db 613 AGAATACAGACTCGCGGTGAAGCTATGCTAACGCCCATGAGGCTGTATCTGAGAGCTT 672
OY 103 LeuLeu11eGlyValHisAspAsp11eValSerMetGluTyrAspLeuAlaTyrLysLeu 122
    ||| ||| ::::: ||| ::::: ||| ::::: |||
Db 673 GCACGTGAAGGCTACTACGACAGCAAGCTCTTGGACGCGGAGCTCCGCTTTGAGATT 732
OY 123 GlYAsp 124
    |||::: |||:::
Db 733 GCGCAG 738

```

RESULT 29
LOCUS BF161010 982 bp mRNA linear EST 30-OCT-2000
DEFINITION 601167637F1 NC1_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3984080 5',
mRNA sequence.

ACCESSION BF161010
VERSION BF161010.1 GI:11041117
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 982)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9185 row: 1 column: 09
High quality sequence stop: 652.

FEATURES
source
Location/Qualifiers
1..982

/organism="Mus musculus"
/strain="Czech 11"
/db_xref="taxon:10090"
/clone="IMAGE:3984080"
/clone_lib="NC1_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
/note="Organ: Lung; Vector: pCMV-SPORT6; Site: 1; Salt:
site: 2; Note: Cloned unidirectionally; Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 325 a 201 c 269 g 187 t

ORIGIN

Alignment Scores:
Pred. No.: 0.401 Length: 982
Score: 89.50 Matches: 26
Percent Similarity: 47.76% Conservative: 38
Best Local Similarity: 19.40% Mismatches: 55
Query Match: 8.38% Indels: 15
DB: 12 Gaps: 3

US-10-024-955-7 (1-213) x BF161010 (1-982)

```

OY 23 AspLys11eThrGluIleAsnLysAlaIleAspAspAlaIle-----IleAla 38
    |||::: |||::: ||| ||| ||| |||
Db 45 GACGAATTCAGTATGATGTGACCAAGCGAGTAAGTACCTGACGACGAGCGAATCTTA 104
OY 39 Ala11eGluInserGIuThr-----11eAspPrometLysValProAspHisAla 55
    ::::: ||| ||| ||| ||| ||| |||
Db 105 GAACAGAGAGAGAGATGAGAGAGCTCAAGTTGAAGTGTCAAGCTGACAGAGATTTC 164
OY 56 AspLysPheGluArgHisValGlyIleValAspPheLysGlyGluLeuAlaMetArgAsn 75
    ||| ||| ||| ||| ||| |||
Db 165 GATATTGCCAAAGACAGATGATTTTGAATTCGACACAGTCACGAGGAGAAAGGA 224
OY 76 11eGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGlu 95
    ::::: ||| ||| ||| ||| |||
Db 225 GTGGAATCCCTCAGAACGAGCTGTGCTGACTCCAGGACACATCTGCAGAAAGCT--- 281
OY 96 GluGlyIleValLysAlaHisLeuLeu11eGlyValHisAspAsp11eValSerMetGlu 115
    |||::: |||::: ||| ||| ||| |||
Db 282 -----CTAATTGCCAATTTGCACCAACATGTTGTCTCTTCA 320
OY 116 TyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrHisVal11eSerAsp11e 135
    ||| ||| ||| ||| ||| |||
Db 321 ATTAGTGAGGCCACCTGCCCTCGGTAACTTAGACTTACGTTACGTTCAACTCCAGAGATG 380
OY 136 GluAspPheValAlaLeuSerLeuGlu11eSerAsp11e 149
    ::::: ||| ||| ||| ||| |||
Db 381 GAAGCTACCAATTTGCCCTTAGAACAGAACTGATGAA 422

```

RESULT 30
LOCUS A2930782 507 bp DNA linear GSS 01-APR-2001
DEFINITION 474.dh258e10.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces
unisporus genomic clone 474.dh258e10.s1, DNA sequence.
ACCESSION A2930782
VERSION A2930782.1 GI:13501692
KEYWORDS GSS.
SOURCE Saccharomyces unisporus.
ORGANISM Saccharomyces unisporus.

REFERENCE 1 (bases 1 to 507)
AUTHORS Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
M.R., Waterston,R.H. and Johnston,M.
TITLE Surveying Saccharomyces genomes to identify functional elements by
comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
Location/Qualifiers
1..507

FEATURES
source
Location/Qualifiers
1..507

/organism="Saccharomyces unisporus"
/strain="NRRL Y-1556 (CBS 396)"
/db_xref="taxon:27294"
/clone="474.dh258e10.s1"
/clone_lib="Saccharomyces unisporus NRRL Y-1556"

Db	204	GCACGACGACACCCCTCCGACACATCCCGCCCAACGCCACACTGTATTGAG-----GTG	257
QY	63	glttlevaaspheylsglycluleuametaragsnilegualaargolyuleuls	82
Db	258	GAGCTGTGTAGTTCAAGAGGAAAGACCTTACAGAGAAGAAAGATGGGGATCATCCGC	317
QY	83	glnmetlysarvglnglyaspalaasnvalylsglygluglnuglytlevallysalahis	102
Db	318	AGAAATACGAGCTCGGGGGAAGGCTATGCCAGGCCCAATGATGCTATGCTGGAACTG	377
QY	103	leuuleleglyvalthiaspaspilaealsermetguitryaspilaeu1atyrileysleu	122
Db	378	GCCCTGGAAAGCTACCAACAGACCGCCTCTTTGACCAAGCGGAGCTTCGTTGAAGTC	437
QY	123	glyasp	124
Db	438	GGGGAC	443
RESULT	27		
LOCUS	AL514856	920 bp	mRNA
DEFINITION	AL514856	LT1_NFL006_PL12	Homo sapiens cDNA clone
ACCESSION	AL514856		prime mRNA sequence.
VERSION	AL514856.1	GI:12778349	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Li, W.-B., Gruber, C., Jesse, J., and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seque@genoscope.cns.fr , Web : www.genoscope.cns.fr .		
FEATURES			
SOURCE	1..920		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CL0BB0142D08"		
	/clone_id="LT1_NFL006_PL12"		
	/issue_type="placement"		
	/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : liangellife@life.com URL : http://fulllength.invitrogen.com "		
BASE COUNT	232 a	229 c	284 g
ORIGIN		170 t	5 others
Alignment Scores:			
Pred. No.:	0.363	Length:	920
Score:	89.50	Matches:	30
Percent Similarity:	45.08%	Conservative:	25
Best local Similarity:	24.59%	Mismatches:	50
Query Match:	8.38%	Indels:	17
Db:	9	Gaps:	3
US-10-024-955-7 (1-213) x	AL514856 (1-920)		
QY	18	asprolethistyraspylserhrglucguileansylalaileaspaalalle	37
Db	349	GACCAATTCCTCCCTGGGAAABAGGCGATCAAGGCTTGGGACATTTCCATA	408

QY	38	AlaAlaIleGluInSerGlnThrIle-----	-----	46
Db	409	GCCACCAATGAAGGTCGGGAGAGGTGTGCCACATCACTCACCTCAAAACCAAGAAATATGCTACAGCT		468
QY	47	-----AspPrometIysValProAspHisAlaAsp---LysPheGluArgHisVal		62
Db	469	TCAGCAGGCAGCTCCCAAAATATTCGCCCAATGCGACAGCTGTGATTTAGAG-----GTG		522
QY	63	GlyIleValAspPheIleGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLys		82
Db	523	GAGTGTGTTGAGTTTAAAGGAGAAATCTGACGGAAGAGGAAGAGCGCGGAATCATTCGC		582
QY	83	GlnMetIysArgGlnGlyAspAlaAsnValIysGlyGluGluGlyIleValIysAlaHis		102
Db	583	AGAATACAGACTCGCGGTGAAGCGCTATCTCTAAGCCCAATGAGGGTGTATCTGTGAGGTT		642
QY	103	LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu		122
Db	643	GCACGTGAAGAGGTACTACAAGACAGACTCTTTGACACGCGGAGACTCCGCTTGAGATT		702
QY	123	GlyAsp		124
Db	703	GGCGAG		708
RESULT 28				
LOCUS	BM461957	929 bp	mRNA	linear
DEFINITION	AGENCOURT_6418161 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5534019			
ACCESSION	BM461957			
VERSION	BM461957.1	GI:18510997		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	1 (bases 1 to 929)			
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: c9apbs@email.nih.gov			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LMNL at:			
	http://image.llnl.gov			
	plate: LM412219 row: n column: 04			
	High quality sequence stop: 610.			
FEATURES	Location/Qualifiers			
Source	1..929			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:5534019"			
	/clone_lib="NIH_MGC_71"			
	/tissue_type="leiomyosarcoma"			
	/lab_host="DH10B (phage-resistant)"			
	/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;			
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.			
	Average insert size 2.1 kb.			
BASE COUNT	210 a 265 c 291 g 163 t			
ORIGIN				
Alignment Scores:				
Pred. NO.:	0.368	Length:	929	
Score:	89.50	Matches:	30	
Percent Similarity:	45.08%	Conservative:	25	
Best Local Similarity:	24.59%	Mismatches:	17	
Query Match:	8.38%	Indels:	3	
DB:	13	Gaps:	3	


```

Db      611 ATCACGAGAAATGATGACTTGCC 640
RESULT 25
BE536812
LOCUS    BE536812                722 bp    mRNA    linear    EST 09-AUG-2000
DEFINITION 60106481EF1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3451010 5',
            mRNA sequence.
ACCESSION BE536812
VERSION   BE536812.1 GI:9765457
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 722)
AUTHORS   NIH-MGC http://mgs.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLM8430 row: f column: 03
           High quality sequence stop: 605.
FEATURES
   source
       1..722
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="IMAGE:3451010"
           /clone_lib="NIH_MGC_10"
           /cell_line="MGC36"
           /lab_host="DH10B"
           /note="Organ: cervix; Vector: PCMV-SPORT6; Site_1: NotI;
           Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
           Average insert size 1.5 kb. Library prepared by Life
           Technologies."
BASE COUNT  210 a 156 c 218 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 0.183      Length: 722
Score: 90.50      Matches: 30
Percent Similarity: 45.90%      Conservative: 26
Best Local Similarity: 24.59%      Mismatches: 49
Query Match: 8.47%      Indels: 17
DB: 10      Gaps: 3
US-10-024-955-7 (1-213) x BE536812 (1-722)
Qy      18 AspProIleHisTyrAspIleThrGluIleAsnLysAlaIleAspAlaIle 37
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      11 GACAAATTCCTCTTGACCTGGGAAAGGAGGATCATCAAGCCTTGACATTCGCATTA 70
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      38 AlaAlaIleGluGlnSerIleThrIle----- 46
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      71 GCCACCATGAGGTGGGGAGGTGCCACATCACCTGCAGAACAGATATATGCTACGGT 130
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      47 -----AspProMetLysValProAspHisAlaAsp---LysPheGluatGHisVal 62
      ||| ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      131 TCAGCAGCGACGTCTCCAAAGATTCGCCCAATGCCACCTGTGATTGAG-----GTG 184
      ||| ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Qy      63 GlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluIleAlaArgGlyLeuLys 82
      ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      185 GAGTTGTAGAGTTTAAGGAGAGATCTGACGAGAGAGAGATGCGGAAATCATTCGC 244
      ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      83 GluMetLysArgGlnGlnLysAlaAsnValLysGlyGluGlnGlyIleValLysAlaHis 102
      ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      245 AGAATACAGACTCGCGGTGAGGCTATGCTAAGCCCAATGAGGTCCTATCGTGAGGTG 304
Qy      103 LeuLeuIleGlyValHisAspAlaIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
      ||| ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      305 GCACGTGGAAGGCTACTACAGACAGACAGCTTTGACACAGGAGGAGCTCCGCTTTGAGATT 364
Qy      123 GlyAsp 124
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      365 GCGCAG 370
RESULT 26
BI104634
LOCUS    BI104634                754 bp    mRNA    linear    EST 26-JUN-2001
DEFINITION 602891509F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5036451 5',
            mRNA sequence.
ACCESSION BI104634
VERSION   BI104634.1 GI:14555527
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 754)
AUTHORS   NIH-MGC http://mgs.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Gilbert Smith, Ph.D.
           cDNA Library Preparation: Life Technologies, Inc.
           DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLM1101 row: b column: 04
           High quality sequence start: 2
           High quality sequence stop: 639.
FEATURES
   source
       1..754
           /organism="Mus musculus"
           /strain="CZECH 17"
           /db_xref="taxon:10090"
           /clone="IMAGE:5036451"
           /clone_lib="NCI_CGAP_Lu29"
           /tissue_type="spontaneous tumor, metastatic to mammary.
           Stem cell origin."
           /lab_host="DH10B"
           /note="Organ: lung; Vector: PCMV-SPORT6; Site_1: SalI;
           Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
           Library constructed by Life Technologies. Investigator
           providing samples: Gilbert Smith, NIH"
BASE COUNT  197 a 179 c 243 g 135 t
ORIGIN
Alignment Scores:
Pred. No.: 0.196      Length: 754
Score: 90.50      Matches: 29
Percent Similarity: 45.08%      Conservative: 26
Best Local Similarity: 23.77%      Mismatches: 50
Query Match: 8.47%      Indels: 17
DB: 13      Gaps: 3
US-10-024-955-7 (1-213) x BI104634 (1-754)
Qy      18 AspProIleHisTyrAspIleThrGluIleAsnLysAlaIleAspAlaIle 37
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      84 GACAAATTCCTCTTGACCTGGGAAAGGAGGATCATCAAGCCTTGAGATATTGCTGTG 143
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Qy      38 AlaAlaIleGluGlnSerIleThrIle----- 46
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      144 GCAACCATGAAAGTGGGGAGGTGCCACATCACCTGCAGGCCAGAAATATGCTATGTC 203
      ||| ::::| | | | | | | | | | | | | | | | | | | | | | | | |

```

```

US-10-024-955-7 (1-213) x BG501777 (1-781)
OY 15 ValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAsp 34
    |||||
DB 153 GTCACTGCGAAGAAAGTGAACAA-----ACACACTCAGTGAATGGCATTTACTGAG 203
OY 35 AspAlaIleAlaIleAlaIleGluInserGluThrIleAspPrometLysValProAspHis 54
    ::|||
DB 204 GAAGCTGATCCCACTTACTCTGTGCAAGTAATTCGCCCTGAGAGGTGTGATCA 263
OY 55 AlaAsp---LysPheGluArgHisValGlyIleVal-----AspPheLysGlyGlu 70
    ::|||
DB 264 ACACAGACTGATGACCAAGCAATGATGATGTGAGAGCGCATATGAAAGGTGAG 323
OY 71 LeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGluMetLysArgGlnGlyAla 90
    ::|||
DB 324 -----GTCATTCATTTGGCATCGTTGGATGGCCCAACAAGGGGATTTC 368
OY 91 AsnValLysGlyGluGluGluIleValLysAlaHisLeuLeuIleGlyValHisAspAsp 110
    |||||
DB 369 CTGCAGAAAGGGAGAGC-----GTCAAGTTCATTCATGTGTGTC---CTGGGCCAAAAAT 419
OY 111 IleValSerMetGluTyrAspLeuAla-----119
    ::|||
DB 420 GCACAACACTATGCTTACACATCACACCCCTGCCAGGGCCACAGTGAATGTGTGAA 479
OY 120 -----TyrLysLeuGlyAspLeuHisProThrThrHisValIle 132
DB 480 GATCAGTTGGCTTCATTACTATGATGAGTAGAGATGACAGAAACTTTTTCCTGAG 539
OY 133 SerAspIleGlnAspPheValValAlaLeuSerLeuGluIleSerAspGlnGlyAsnIle 152
    ::|||
DB 540 AAGAAGATTCAGCAT-----GGCATTTGAGCTACAGCAGCAAGATGACGGTGTGAGTT 590
OY 153 ThrMetThrSer 156
    |||
DB 591 CTCAGTGATTC 602

RESULT 24
LOCUS AI297539 653 bp mRNA linear EST 23-APR-2001
DEFINITION LP11861.5prine LP Drosophila melanogaster larval-early pupal
Drosophila melanogaster cDNA LP11861 5 similar to
BG:DS00941.12: FBan0008997 located on: 2L 34D4-34D4: 04/11/2001,
mRNA sequence.
ACCESSION AI297539
VERSION AI297539.1 GI:3946946
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 653)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMT Drosophila EST Project
TITLE Unpublished (2001)
JOURNAL
COMMENT Other_ESTs: LP11861.3prine
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, estefruitfly.berkeley.edu
hit genomic sequence AC001659
Plate: LP.118 row: F column: 1
High quality sequence stop: 545
POLKA-No.
FEATURES
source 1..653
/organism="Drosophila melanogaster"

```

```

/dB_xref="taxon:7227"
/clone="LP11861"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pot2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: pOT2; Site-1: EcorI;
Site-2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

BASE COUNT 162 a 155 c 178 g 157 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.157 length: 653
Score: 90.50 Matches: 56
Percent Similarity: 38.00% Conservative: 39
Best Local Similarity: 22.40% Mismatches: 58
Query Match: 8.47% Indels: 97
DB: 9 Gaps: 14

US-10-024-955-7 (1-213) x AI297539 (1-653)
OY 1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAla----- 17
    ::|||
DB 41 ATAAATGAA---ATCTTCGTGCTATTGTGGCTTTGTGGCGGTTCGACGCGCTTCC 97
OY 18 -----AspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAsp 35
    |||||
DB 98 ATGGGCCACCCCATGAA-----ACCACTGCATCTCCAGCAGCATTTGTGAT 145
OY 36 AlaIleAlaIleGluInser-----GluThr 45
    |||
DB 146 GTGATCGAGCGATTAGAGACAGATGCCCTGTGCTTACCAGTGTGGCTGCCGCCA 205
OY 46 IleAspPrometLysValProAspHisAlaAspLysPheGluArgHisValGlyIleVal 65
    ::|||
DB 206 TTGGCTCTCTTGAATTC---GACCATCAGAT-----235
OY 66 AspPheLysGlyLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLys 85
    |||||
DB 236 -----ATCAACATC-----244
OY 86 ArgGlnGlyAspAlaAsnValLysGlyGluGluIleValLysAlaHisLeuLeuIle 105
    ::|||
DB 245 -----GATAGCAGCGTGTGAAGGCTCGAGGAACCATGATTCATTCGCGTGAAC 295
OY 106 GlyValHisAsp---AspIle-----111
    |||||
DB 296 GGTCTGAATGATTTGATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 355
OY 112 -----ValSerMetGluTyrAspLeuAlaTyrLysLeu 122
    ::|||
DB 356 TACAAGTTCATTTCCGCGCAGCTGAAGCTGATGATGATGATGATGATGATGATGATGAT 415
OY 123 GlyAspLeuHisProThrThrHisValIleSer-----AspIleGln 136
    ::|||
DB 416 AAGAAGTACAGGATTTACATCAATCTGATCGGTGCCCGCAATGCCAATTCGCATCAAG 475
OY 137 AspPheValValAlaLeuSerLeuGluIleSer-----AspGlnGlyAsnIleThr 153
    |||||
DB 476 GATATGCTTATTCGGGGCAGTATGAAAGTCTCACTCGGGGTGATGAGCGCAACGTGAAG 535
OY 154 MetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyLeuSer 173
    ::|||
DB 536 CTGAAGTTCATTCGAGGTCCGC-----ACCCACTTGCGGT-----GAG 571
OY 174 IleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnSptThr 193
    ::|||
DB 572 GTTGATTCGAGATCGAGGCGATCTCCGNGCAT-----GGCAGC 610
OY 194 ValArgGlyGluMetThrLysValLeuAla 203

```

12	Gaps:	8
----	-------	---

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LML at:
<http://image.lnl.gov>
 Plate: LCM2571 row: d column: 17
 High quality sequence stop: 608.

FEATURES

source

1..944

/organism="Homo sapiens"
 /db.xref="taxon:9606"
 /clone="IMAGE:6381280"
 /clone_lib="NIH_MGC_40"
 /tissue_type="Carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
 Site:2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(G). Library constructed by
 Ling Hong in the Laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: This is a NIH_MGC Library."

BASE COUNT 265 a 201 c 268 g 205 t 5 others
 ORIGIN

Alignment Scores:

Pred. No.: 0.203 Length: 944
 Score: 91.50 Matches: 30
 Percent Similarity: 45.08% Conservative: 25
 Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.57% Indels: 17
 Gaps: 3

US-10-024-955-7 (1-213) x BG943606 (1-944)

QY 18 ASPPTOLIEHSTYRASPlystleThrgluciluileasnlyAlaileaspAlaile 37
 Db 142 GACAAATTCCTCTTGACCTGNGAAGGAGGATCATCAAGGCTTGCGACATTCGCATA 201
 QY 38 AlaAlaileglugInserGlu----- 44
 Db 202 GCCACCATGAAGGTGGGGAGGTGTGCCACATCACCCTCAACCAAGAAATATGCCCTACGCT 261
 QY 45 ---ThrlleasprPromelysValProasphisaAlaasp---LysPheglunarghisVal 62
 Db 262 TCAGCAGCAGCTCCCTCAAGATTCCTCCCAATGCCAGCTGTGTTCAG-----GTG 315
 QY 63 GlyIleValaspPheIysGlyGluLeuAlaMetargasnIleGluAlaArgGlyLeuLys 82
 Db 316 GAGTTGTTGAGTTTAAGGAGAGATCTGACGAAAGAGAAATGCGCGAATCATTCGC 375
 QY 83 GlmetIysArgGlnGlyAspAlaAsnValIysGlyGluGluGlyIleValIysAlaHis 102
 Db 376 AGAATACAGACTCGCGGAGAGGCTATGCTTAAGCCCAATGCGATCTCGAGAGCTT 435
 QY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaIatyrLeu 122
 Db 436 GCACCTGGAGGTTACTACAGACAGACTCTTTGACACAGCGGAGACTCCGCTTTGAGATT 495
 QY 123 Glyasp 124
 Db 496 GCGCAG 501
 RESULT 21
 BE583938/c BE583938 673 bp mRNA linear EST 16-AUG-2000
 LOCUS 6-7H-HA PsojleHA Glycine max/Phytophthora sojae mixed EST library
 DEFINITION cDNA, mRNA sequence.
 ACCESSION BE583938
 VERSION BE583938.1 GI:9834887
 KEYWORDS EST.
 SOURCE Glycine max/Phytophthora sojae mixed EST library.
 ORGANISM Glycine max/Phytophthora sojae mixed EST library

REFERENCE

1 (bases 1 to 673)

AUTHORS Oubr.D., Hraber,P.T., Sobral,B.W.S. and Gijzen,M.

TITLE Comparative analysis of expressed sequences in phytophthora sojae

JOURNAL Plant Physiol. 123 (1), 243-254 (2000)

MEDLINE

20267956

COMMENT

Contact: Gijzen M
 Agriculture and Agri-Food Canada
 1391 Sandford Street, London, Ontario, Canada N5V 4T3
 Tel: 519 457 1470
 Fax: 519 457 3997
 Email: gijzenm@em.agr.ca
 Transcript obtained from mixed plant-pathogen interaction culture.
 Location/Qualifiers

FEATURES

source

1..673

/organism="Glycine max/Phytophthora sojae mixed EST
 library"
 /strain="Phytophthora sojae race 2 strain P6497"
 /cultivar="Glycine max cultivar Harosoy"
 /db.xref="taxon:135715"
 /clone_lib="PsojleHA"
 /tissue_type="Plant hypocotyls infected with Phytophthora
 sojae"

/dev_stage="8 d old etiolated hypocotyls 48 h
 post-inoculation with phytophthora sojae zoospores"
 /lab_host="E. coli strain XL0R"

/note="Vector: pBK-CMV; Site:1: EcoRI; Site:2: XhoI; This
 cDNA library was constructed from polyA+ enriched mRNA
 from etiolated hypocotyls 48 h post-inoculation with
 Phytophthora sojae zoospores. Complementary DNA was
 synthesized from mRNA using an XhoI-poly(dT)
 linker-primer. EcoRI adapters were ligated to the
 blunt-ended cDNA fragments and the products were digested
 with XhoI for directional cloning into lambda ZAP Express
 vector. This lambda library was amplified once using E.
 coli host strain XL1 Blue MR+. Inserts were then
 subcloned by mass excision using Exsist helper phage
 for conversion into phagemid vector pBK-CMV in E. coli
 host strain XL0R. Sequenced using T3 primer: 5' ATT AAC
 CCT CAC TAA AGG GA 3'."

BASE COUNT 141 a 202 c 200 g 129 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 0.141 Length: 673
 Score: 91.00 Matches: 29
 Percent Similarity: 49.45% Conservative: 16
 Best Local Similarity: 31.87% Mismatches: 34
 Query Match: 8.52% Indels: 12
 Gaps: 3

US-10-024-955-7 (1-213) x BE583938 (1-673)

QY 28 GluIleAsnLysAlaIleAspAspAlaIleAlaIleGluGlnInserGluThrlleasp 47
 Db 331 GAGGTGAAGAAAGCGGTAGACAGCCAGTAGT-----CAG 296
 QY 48 PrometIysValProasphisaA-----AspIysPheglunarghisValIyle 64
 Db 295 CCACGTGGGTGTCGTCGCCCGCGGCTGTGATGCTTGAACACATGTAGGCCGG 236
 QY 65 ValaspPheIysGlyGluLeuAlaMetargasnIleGluAlaArgGlyLeuLysGlnMet 84
 Db 235 GTAGATGAAGCCACAGAGCCACACAGACTCTCGCGCCACACAGAAAGAGAGAT 176
 QY 85 LysArgGlnGlyAspAlaAsnValIysGlyGluGluGlyIleValIysAlaHisLeu 104
 Db 175 GCCCGCCAGAGCGCGCCACAGACAGATGTAGAA---CTTGTCCAGCCCGTGTGCTT 119
 QY 105 IleGlyValHisAspAspIleValSerMetGlu 115
 Db 118 CTCGAGCTCCAGATCTCTGTAGCGCTCGAG 86

QY	137	AspPheValAlaAlaLeuSerLeuCluIleSerAspGluGluGlyAsnIleThrMetThrSer	156
Db	411	GACTCTTGATCAATCTT-----ATTACTACACCGGGACAGTTGATTTCAGTTCC	461
QY	157	PheGluValATGAGlnPheAlaAsnValAlaAsnHisIleGlyGlyLeuSerIleLeuAsp	176
Db	462	---GAGGTTACTGCGCGCACTCCGTGTGACCAT-----GTTGCCCTTAGTGCTCTTAA	512
QY	177	ProIlePheGlyVal	181
Db	513	TGTGTGAGAGGAGTGTG 527	
RESULT 18			
LOCUS	TA319G090/c	560 bp	DNA
DEFINITION	T. brucei sheared genomic DNA clone 319g09, reverse sequence,		linear GSS 13-DEC-2000
ACCESSION	AL492823		
VERSION	AL492823.1	GI:11867593	
KEYWORDS	GSS.		
SOURCE	Trypanosoma brucei.		
ORGANISM	Trypanosoma brucei Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		
REFERENCE	1 (bases 1 to 560)		
AUTHORS	Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.		
COMMENT	Direct Submission Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TRBU927/4 G9rat 10.1) was mechanically sheared to give a light size distribution (4 kb). The v + 1 method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999). Email: nelsayed@tigr.org Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/T_brucei/ Location/Qualifiers		
FEATURES	1..560		
SOURCE	/organism="Trypanosoma brucei" /strain="TRBU927" /db_xref="taxon:5691" /clone="319g09"		
BASE COUNT	136 a 148 c 154 g 122 t		
ORIGIN			
Alignment Scores:			
Pred. NO.:	0.057	Length:	560
Score:	93.00	Matches:	44
Percent Similarity:	42.08%	Conservative:	33
Best Local Similarity:	24.04%	Mismatches:	66
Query Match:	8.71%	Indels:	40
DB:	17	Gaps:	10
US-10-024-955-7 (1-213) x TA319G090 (1-560)			
QY	29	IleAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspPro	48
Db	549	GTCACATTCACCGTGTGAGAGTCCGTCCTCATGAGCACTCCCGACAGATCCGTAAT	490
QY	49	MetLysValProAspHisAlaAspLysPheGluATGHisValGlyIleValAspPheLys	68
Db	489	ATGTCCCGATGCCGACAGCTGCAC-----CACGGCAAG	457

Db	147	GCTCGACTGAGCGTATTCGCCCC	124
RESULT 14			
LOCUS	A0651517	625 bp	DNA linear GSS 22-JUN-1999
DEFINITION	Sheared DNA-8117.TR Sheared DNA Trypanosoma brucei genomic clone		
ACCESSION	A0651517		
VERSION	A0651517.1	GI:5144703	
KEYWORDS	GSS.		
SOURCE ORGANISM	Trypanosoma brucei. Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;		
REFERENCE	1 (bases 1 to 625) El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Sub,E., Malek,J., Fujii,C., Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M., Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library Unpublished (1999)		
JOURNAL COMMENT	Other_GSSs: Sheared DNA-8117.TF Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: nelsayed@tigr.org Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tldb/mdb/tbdb/ . Seq primer: M13-Reverse Class: shotgun.		
FEATURES			
Source	Location/Qualifiers 1..625 /organism="Trypanosoma brucei" /strain="TREU927/4 GUTat 10.1" /db_xref="taxon:5691" /clone="Sheared DNA-8117" /clone_1bp="Sheared DNA" /note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."		
BASE COUNT	163 a 153 c 183 g 126 t		
ORIGIN			
Alignment Scores:			
Pred. NO.:	0.0229	Length:	625
Score:	96.50	Matches:	43
Percent Similarity:	42.86%	Conservative:	32
Best Local Similarity:	24.57%	Mismatch:	65
Query Match:	9.04%	Indels:	35
DB:	17	Gaps:	9
US-10-024-955-7 (1-213) x A0651517 (1-625)			
OY	24	Lysilrhrngluaiilaasnllysalaleaspapalailelaialeleuglnser	43
Db	462	AAGCATCAAGATGCACAACCTCACCGTTGTAGTGAGTCCGGCCTCATGACCTAACCCC	403
OY	- 44	Giurrrileapromellryvalproasphisaiaaplysphegluarghisvalgly	63
Db	402	GAAACAGATCCGTAAATATCTCCGTGATWCGCGACGTCGAC-----	364

Oy	64	IleValAspPheIlysgLyluleLeuAlaMetKraYrpsnIlleglAlaArGlyLeuysLn	83
Dd	363	-----CACGGCAAGTCCACCTGTGTGCACATCGCTTGTCGGTGGCCGCTGTATCATTTAAG	310
Oy	84	MetytAsrGngInglYasPalaasnVal-----LysGlYlgUgUglYlLeVallys	100
Dd	309	ATGAGAGACGCCGGTGAACAACGTATCATGTGACACTCGTGGGACAGAA---ATTGGCCGT	253
Oy	101	AlaHisLeuLeuIlleglYvalHIsAspAspileValSerMetGUtyrAspLeuAlArYr	120
Dd	252	GGT-----ATCACAATCAAGTCGATCGATCCATCTGTGATGACATTAC-----	214
Oy	121	LysLeuGlYasrPLeuHIsProthrHrhisValIlleserAspile-----Gln	136
Dd	213	-----CANGTACCTCCCAGATATTATCATGTACCCTTCGCCGATGACGCCGC	169
Oy	137	AspHeValValAlaLeuSerLeuclulIleserAsprGlYasIleHrmetHrsr	156
Dd	168	GACTTCGTGATCAATCT-----ATTGACTCACCCGGGACAGTGTAATTCAATTCC	118
Oy	157	PheglUValArGlnPhePalasNValYalsnHisIlleglYglYLeuSerlleuAsp	176
Dd	117	--GAGCTTACTGTGCATCGCTCGGTGACCGAT-----GGTGCCCTTAGTGCTGTGAT	67
Oy	177	ProIlePheglYvalLeuSerAspValLeuHraIllephegn	191
Dd	66	TGTGTGAGGGGTGTGTGCTGCATAACGGAACGGTGTCTTCCGACG	22
RESULT	15		
LOCUS	BM276304	523 bp	mRNA linear EST 20-DEC-2001
DEFINITION	PFSTroaa1b01.y1 Plasmodium falciiparum 3D7 gametocyte cDNA library		
ACCESSION	Plasmodium falciiparum 3D7 cDNA 5' mRNA sequence.		
VERSION	BM276304		
KEYWORDS	BM276304.1 GI:17969658		
SOURCE	EST.		
ORGANISM	Plasmodium falciiparum 3D7.		
REFERENCE	Plasmodium falciiparum 3D7.		
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
JOURNAL	1 (bases 1 to 523)		
COMMENT	Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Page,D., Matta,M., Hillier,L., Martin,J., Wylie,T., Dente,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagarashvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Watkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley.D. Washu Plasmodium Est Project unpublished (2001) Contact: L. David Sibley Washu Plasmodium Est Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu Library was constructed by R. Haywood. DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University Seq primer: -40UP from Glpco High quality sequence stop: 428. Location/Qualifiers 1..523 /organism="Plasmodium falciiparum 3D7" /db_xref="taxon:36329" /clone_id="Plasmodium falciiparum 3D7 gametocyte cDNA library" /dev_stage="gametocyte (stage III-V)" /lab_host="DH10B (Genetho, Invitrogen, Inc.)" /note="vector: pBluescript SK plus; Site.1: EcoRI; Site.2: XhoI; The library was constructed by R Haywood. cDNAs were synthesized from gametocyte poly(A)+ RNA by oligo d(T)		

KEYWORDS EST.
cotton bollworm.
SOURCE Helicoverpa armigera
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
REFERENCE 1 (bases 1 to 673)
AUTHORS Grubor,V., Kuczek,E., Wilson,P. and Heckel,D.G.
TITLE Construction and analysis of a cDNA library from larval midguts of cotton bollworm *Helicoverpa armigera*
JOURNAL Unpublished (2002)
COMMENT Contact: Vladimir Grubor
CESAR-Centre for Environmental Stress and Adaptation Research
Department of Genetics, The University of Melbourne
Parkville, Victoria, 3010, Australia
Tel: +61 3 8344 6246
Fax: +61 3 8344 5139
Email: vgrubor@pgrad.unimelb.edu.au
Seq primer: T3 Forward.
FEATURES
Location/Qualifiers
source
1..673
/organism="Helicoverpa armigera"
/strain="AN02"
/db_xref="taxon:29058"
/clone_lib="Helicoverpa armigera larval midgut cDNA library"
/tissue_type="Midgut"
/dev_stage="Fifth instar larvae"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript II SK(-); Total RNA was prepared from midguts of mid-fifth instar larvae of *Helicoverpa armigera* using the RNagents kit (Promega). PolyA mRNA was obtained using the Dynabeads mRNA purification kit (Dynal). First-strand cDNA was made by oligo dT-priming with 5'-(GA
10'-ACTAGTCGACGTTTCTTTT-3'. Following second-strand cDNA synthesis, size-selected products >600 bp were ligated to 5' linkers 5'-AATTCGACGACG-3' and 5'-CCTCGTCCG-3', directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites, and packaged using GigaPack III Gold Packaging Extracts. The library was amplified, and plasmids were mass-excised using ExAssist helper phage (Stratagene) and propagated in XLI-Blue MRF cells. Single-pass 5' end sequencing was performed using the T3 primer. Library construction supported by CESAR, a Special Research Centre of the Australian Research Council. Sequencing supported by the Director's Discretionary Fund of the Australian Genome Research Facility."
BASE COUNT 175 a 157 c 166 g 175 t
ORIGIN
Alignment Scores:
Pred. No.: 0.00295 Length: 673
Score: 103.50 Matches: 38
Percent Similarity: 42.35% Conservative: 34
Best Local Similarity: 22.35% Mismatches: 91
Query Match: 9.69% Indels: 7
DB: 14 Gaps: 4
US-10-024-955-7 (1-213) x BU038592 (1-673)
QY 29 IleaenlyalaIleasp-----AspAlaIleAlaIleGlucInserGIuThr 45
DB 509 GTCAACGACTATGTGACATACCATTCGATCGTCCCTTCATCCGAAGCATGGC 450
QY 46 IleaSPromelyValrProasPhisAlaasplysPheGIuAghIsvAlaGIleVal 65
DB 449 CTGATTCCAATGACGTCCGCGACGTCATCGAAGGTTTGAAGTCGACGCTTGGCTGATC 390
QY 66 AspPheLySGlyGlueuAlaMetArGasnIleGluaIaArgGIleuLySGlnMetLys 85
DB 389 ACGTATAGCGCGTGGCTGAATACACGATGTTACATGACGTGCTTGTGGAACGTGTC 330

QY 86 ArgGInGIyAspAlaAsnValySGlyGIuGIuGIyIleValIyAlaHisLeuLeu 105
DB 329 CCCTCGGGGACACAGAACTTAACACTTCGCCAGATTTGCGTGTGGCTGCACAG 270
QY 106 GIyValHisAspAspIleValSerMetGIuYrAspLeuAlaTyrlsLeuGIyAspLeu 125
DB 269 CAATTCCACCGAT-----CTTAAGTTTCATCTCAAAATACCTGGTGAAGTCATGAACCTTG 216
QY 126 HisProthThHisValIleSerAspIleGIuAspPheValValAlaLeuSerLeuGIu 145
DB 215 GGACSTACAGGTGGATTCATCGCAACCCGCTTCGGTTATGCTCATGTATT 156
QY 146 IleaSer---AspGIuGIyAsnIleThrMetThrSerPheGIuValArgIInPheAlaAsn 164
DB 155 ATTGACTTCAACAATGACAGATTCACCTTCACAGTTACAGCTCATGACATCGGCCGT 96
QY 165 ValValAsnHisIleGIyGIleuSerIleLeuAspProIlePheGIyValLeuSerAsp 184
DB 95 CTTCGCGTTGCGATTCACGTGCAACATTTTGACTGACGTGCTGTGAACCCGCTTCACAC 36
QY 185 ValLeuThAlaIlePheGIuAspThrVal 194
DB 35 GTGTTCGTCAGAATCTTC--GACACCATC 9
RESULT 12
TAB6F08P
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 86f08, forward sequence,
genomic survey sequence.
ACCESSION
AL462038
VERSION
AL462038.1 GI:11862624
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei.
ORGANISM
Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 537)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
JOURNAL
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nilesanger@ac.uk
COMMENT
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + l method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nilesayer@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at <http://www.sanger.ac.uk/Projects/T-brucei/>.
FEATURES
source
1..537
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="86f08"
BASE COUNT 112 a 154 c 130 g 140 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 0.00614 Length: 537
Score: 100.00 Matches: 45
Percent Similarity: 43.09% Conservative: 36
Best Local Similarity: 23.94% Mismatches: 67
Query Match: 9.36% Indels: 40

```

Qy 20 llehstYrAspLysIleThrgluIleAsnLysAlaIleAspAlaIleAla 39
Db 185 -----GATCGATCAATAAATTTGGCGCATTCATATATGGCAAT 223
Qy 40 llegluInserGluThrIleAspPrometLysValProAspHisAlaAsp----- 56
Db 224 TTAATCAAAATCAAAAGAAATGGAATCCATTAATTT-----CATCAAGATTATTAATTTA 277
Qy 57 ---LysPheGluArgHisValGlyIleValAspPheLysGlyLeuIleAlaMetArgAsn 75
Db 278 GAAGTCAATGAAAAAATATTGGACCATTCACATACATGGTACGGTTAAATATCCAA 337
Qy 76 llegluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGly 95
Db 338 GGTGAATGTGTGGCTTAACTTAAAGACGTGTGGTATGTAATGACCAATGAT 397
Qy 96 GUGLY-----lleValLysAlaHisLeuLeuIleGlyValHisAspAlleVal 112
Db 398 GAATCGGCAATATGTGTGGAAGTTGACCTTATA-----ACACGTAAAGAAATCATTA 451
Qy 113 SerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisPro-----ThrThrHis 130
Db 452 ---TTGATGATCGCATTTTACTGCAAAATATTGGTACATACCATTCATTCGCCATCCG 508
Qy 131 ValIleSerAspIleGlnAspPheValAlaAlaLeuSerLeuGluIleSerAspGly 150
Db 509 ATGTGTTGCCACATCAGAGTATGATTTATTCCTTGCTGGTTAAAGTGTGCA----- 562
Qy 151 AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnVal 165
Db 563 AATTAAACGCTTACCAATTTGTTACGAATGAAATTCGATGTT 607
RESULT 10
LOCUS B0834844 630 bp mRNA linear EST 08-AUG-2002
DEFINITION Po.ad_04B03_TEXF1 Psoroptes ovis mixed Psoroptes ovis cDNA clone
ACCESSION B0834844
VERSION B0834844.1 GI:22139158
KEYWORDS EST.
SOURCE sheep scab mite.
ORGANISM Psoroptes ovis
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Sarcopitoidae; Psoroptidae;
Psoroptes.
REFERENCE 1 (bases 1 to 630)
AUTHORS Kenyon,F., Welsh,M., Parkinson,J., Whitton,C., Blaxter,M. and Knox
,D.P.
TITLE Expressed sequence tag survey of gene expression in the scab mite
Psoroptes ovis - allergens, proteinases and free radical scavengers
JOURNAL Unpublished (2002)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JF, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared for Dr David Knox, Moredun Research
Institute, UK Sequencing was performed by Mark Welsh, ICAP,
Edinburgh
PCR Primers
FORWARD: M13R
BACKWARD: T7PL
Plate: 04 Row: B Column: 03
High quality sequence stop: 360.
FEATURES
source
1..630
Location/Qualifiers
/organism="Psoroptes ovis"
/db_xref="taxon:83912"
/clone="Po.ad_04B03"
/clone_11b="Psoroptes ovis mixed"

```

```

/sex="Mixed"
/deb_stage="Mixed (predominantly adult)"
/Note="Vector: Lambda Triplex2; Site 1: SfilA; Site 2:
SfilB; Psoroptes ovis is the infective agent causing sheep
scab, a serious dermatitis in sheep. The cDNA was
obtained from mites harvested from the skin of previously
infected sheep harbouring a moderate challenge
infection. The cDNA was cloned using the Clontech Lambda
Triplex2 Smart cDNA cloning system."
BASE COUNT 202 a 90 c 123 g 215 t
ORIGIN
Alignment Scores:
Pred. No.: 5,58e-05 Length: 630
Score: 116.00 Matches: 49
Percent Similarity: 45.32% Conservative: 43
Best Local Similarity: 24.14% Mismatches: 80
Query Match: 10.86% Indels: 31
DB: 14 Gaps: 11
US-10-024-955-7 (1-213) x B0834844 (1-630)
Qy 1 MetMetLysPhe---LeuLeuIleAlaIleAlaPheValAlaValSerAlaAspPro 19
Db 68 ATGATCAAAATTTGCTGATTTGTTGGTTCGTTACACTTGTCTCCTCGCAATGACCCA 127
Qy 20 llehstYrAspLysIleThrgluIleAsnLysAlaIleAspAlaIleAla 39
Db 128 -----GATCGATCAATAAATTTGGCGCATTCATATATGGCAAT 166
Qy 40 llegluInserGluThrIleAspPrometLysValProAspHisAlaAsp----- 56
Db 167 TTAATCAAAATCAAAAGAAATGGAATCCATTAATTT-----CATCAAGATTATTAATTTA 220
Qy 57 ---LysPheGluArgHisValGlyIleValAspPheLysGlyLeuIleAlaMetArgAsn 75
Db 221 GAAGTCAATGAAAAAATATTGGACCATTCACATACGTTACGGTTAAATATCCAA 280
Qy 76 llegluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGly 95
Db 281 GGTGAATGTGTGGCTTAACTTAAAGACGTGTGGTATGTAATGACCAATGAT 340
Qy 96 GLY-----GlyIleVal-LysAlaHisLeuLeuIleGlyValHisAspAlleValSe 113
Db 341 GAATCGGCAATGTGTGGAAGTTGACCTTATA-----ACACGTAAAGAAATCATTA-- 392
Qy 113 rMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisPro-----ThrThrHisVa 131
Db 393 -TTAGATGCCGATTTTACTGCAAAATATTGGTACATACCATTCATTCGCCATTCGGAT 451
Qy 131 lleSerAspIleGlnAspPheValAlaAlaLeuSerLeuGluIleSerAspGly 151
Db 452 TGTGCCACATCAGAGTATGAAATTTATCCTTGCGTTAAAGTGTGCA-----AM 505
Qy 151 nIleThrMetThrSerPheGluValArgGlnPheAlaAsnVal-----ValAsnH1 168
Db 506 TTTAAGCCTTACCGAATTTGTTACGAATGAAATTTGCGATGTAACGCTAAAGA 565
Qy 168 sIleGlyGlyLeuSerIleLeuAspProIlePheGlyValLeuSerAspValleuThrAl 188
Db 566 TGCATTTCAATTTGAAGCATTTGATCGAATTTTC-----GATCGATTGACTGC 613
Qy 188 aIlePhe 190
Db 614 ATGCTTT 620
RESULT 11
LOCUS B0038592 673 bp mRNA linear EST 23-AUG-2002
DEFINITION DH01A03 Helicoverpa armigera larval midgut cDNA library Helicoverpa
armigera cDNA 5', mRNA sequence.
ACCESSION B0038592
VERSION B0038592.1 GI:22474116

```


Db	557	AATTAAACGCTTACGCAATTTGTTACGAATGAATAATTCGGATGTTGAATACGGCTA	616
Qy	168	HistHeglyGlyLeuSerIleLeuAspProIleHeglyValLeuSerAspValLeuThr	187
Db	617	GATGCATTTTCATTTGTAAGACATTTGATGCAATTTTC-----GATCGATGACT	664
Qy	188	AlaIlePhe	190
Db	665	GCATGCTTT	673
RESULT 7			
LOCUS	B0834827		
DEFINITION	B0834827	604 bp	mRNA
ACCESSION	B0834827		
VERSION	B0834827		
KEYWORDS	B0834827.1	GI:22139141	
SOURCE	EST.		
ORGANISM	sheep scab mite.		
REFERENCE	1 (bases 1 to 604)		
AUTHORS	Kenyon,F., Welsh,M., Parkinson,J., Whilton,C., Blaxter,M. and Knox		
TITLE	Expressed sequence tag survey of gene expression in the scab mite		
JOURNAL	Psoroptes ovis - allergens, proteinases and free radical scavengers		
COMMENT	Unpublished (2002)		
	Contact: Blaxter ML		
	Institute of Cell, Animal and Population Biology		
	University of Edinburgh		
	Asworth labs, King's Buildings, West Mains Road, Edinburgh, EH9		
	3JF, UK.		
	Tel: +44 131 650 6760		
	Fax: +44 131 670 5450		
	Email: mark.blaxter@ed.ac.uk		
	The library was prepared for Dr David Knox, Moredun Research		
	Institute, UK Sequencing was performed by Mark Welsh, IACAB,		
	Edinburgh		
	PCR primers		
	FORWARD: M13R		
	BACKWARD: T7PL		
	Plate: 03 row: H column: 07		
	High quality sequence stop: 543.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..604		
	/organism="Psoroptes ovis"		
	/db_xref="taxon:83912"		
	/clone="Po.ad.03H07"		
	/clone_lib="Psoroptes ovis mixed"		
	/sex="Mixed"		
	/dex_stage="Mixed (predominantly adult)"		
	/note="Vector: Lambda Triplex2, site_1: SfilA, site_2:		
	SfilB; Psoroptes ovis is the infective agent causing sheep		
	scab, a serious dermatitis in sheep. The cDNA was		
	obtained from mites harvested from the skin of previously		
	unexposed sheep harbouring a moderate challenge		
	infection. The cDNA was cloned using the Clontech Lambda		
	Triplex2 Smart cDNA cloning system."		
BASE COUNT	191 a	90 c	123 g
ORIGIN			200 t
Alignment Scores:			
Pred. No.:	9.55e-06	Length:	604
Score:	121.50	Matches:	45
Percent Similarity:	44.39%	Conservative:	46
Best Local Similarity:	21.95%	Mismatches:	87
Query Match:	11.38%	Indels:	27
DB:	14	Gaps:	10

OY	3	LysPhe-----LeuleuLlealalaValalaphelvalAlavalserAlasPrroIlehis	21
Db	25	AAATTTGCGTGAATGTGGTTTGTTGGTTACCTGTGCACCTGTCACATGACCACA-----	78
OY	22	TyrAspLysIleThGluGlulIealsuLySaLaIleaSpaPaLaIlelaIaIleleGu	41
Db	79	-----GATGCAGTCATTAATAATTGGCCGATTCATTAATGTGGCAATTTTATTC	123
OY	42	GlnSerGluThrLleaspProMetLysValProAspHisAlaSP-----Lys	57
Db	124	AAAAAAGAATGAATCACCACCTTAAATTT-----CATCAAGATTATTAATTTAGAACGTG	177
OY	58	PheGluArgHISValGLILleValAspPheLysGlyGluLeuAlaMetArgsnlleGu	77
Db	178	AATGAAAAAATATTTGGGACCATTCGCCATACATGTCACGTTAAAATATCCAAAGGTGA	237
OY	78	AlaArgGlyLeuLysGIleMetLysArgInglYaspaLaasnValLysGlyGluGlugly	97
Db	238	TTGTTGGGCGCTTAATAGTTTAAGACGTGTGTGTGATGTGAATAGCACCAATGATGATCG	297
OY	98	-----lLeValLysAlaHisLeuleuilegIyValIHIsAspAspIleValSermet	114
Db	298	GGCATATGATGTTGCTGGAAGTAGTGCACCTTATA-----ACACGTAAAGAATATATA---TTA	348
OY	115	GluTyrAspLeuAlaTyLysLueGlsrLynHisPro-----ThrThHisValille	132
Db	349	GATGCTGATTTTATAGCGCAAATATTTGGTAACTACATTCCTCATTTGCCATTTCCATTTGTT	408
OY	133	SerAspIleGlnAspPheValIValaIleuSerLeuGlnIleSerAspGluGlyAsnIle	152
Db	409	GCCCCATCAGACAGATGATCAATTTTATCCTTGCTGGTTAAAGTTGGTCAA-----AATTTA	462
OY	133	ThreThrSerPheGluValaIArgIlnPheAlaSnVal-----ValAsnHisIle	169
Db	463	ACGCTTACCAGATTTGCTACGAAATCAAAATTTCCGGTGTGTAAGAAATACGGCTTAAAGATGCA	522
OY	170	GlyGlyLysSerIleLeuAspProIlePheGlyValLeuSerAspValleuThrAlaIle	189
Db	523	TTTTCAATTTGAAGACCTTTTGATCGAATTTTCCATGCATTTGACTGCATGGTTTGTAAAGTT	582
OY	190	PheGlnAspThrVal	194
Db	583	TTTCAMTGAACAAGTGT	597
RESULT 8	B0834612	645 bp mRNA linear EST 08-AUG-2002	
LOCUS	B0834612		
DEFINITION	Po.ad.O1B08.TEXE1 Psoroptes ovis mixed Psoroptes ovis cDNA clone Po.ad.O1B08 5', mRNA sequence.		
ACCESSION	B0834612		
VERSION	B0834612.1	GI:22138926	
KEYWORDS	EST.		
SOURCE	sheep scab mite.		
ORGANISM	Psoroptes ovis		
REFERENCE	Eukaryotes; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Acariformes; Sarcopitiformes; Astigmata; Sarcoptoidae; Psoroptidae; Psoroptes.		
AUTHORS	1 (bases 1 to 645) Keynon,F., Welsh,M., Parkinson,J., Whitton,C., Blaxter,M. and Knox,D.P.		
TITLE	Expressed sequence tag survey of gene expression in the scab mite Psoroptes ovis - allergens, proteinases and free radical scavengers unpublished (2002)		
JOURNAL COMMENT	Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK. Tel: +44 131 650 6760 Fax: +44 131 670 5450 Email: mark.blaxter@ed.ac.uk The library was prepared for Dr David Knox, Mordeun Research		

Query Match: 12.41% Indels: 41
DB: 13 Gaps: 7

US-10-024-955-7 (1-213) x BM289894 (1-649)

```

OY 5 Leuileu1leAla1aVala1aPheVala1aVala1aSerAlaAspPro1leHis-----Tyr 22
      ::::::::::::::::::::
DB 62 GTCCGTCGTCCTGCTGCGCTGCTGCGACGACGCGCTCAAGCCTTACAGCCTACGCGTG 121
OY 23 Asplys1leThrGluGlu-----1leAsn1yAla 32
      ::::::::::::::::::::
DB 122 GAAGGGGTGACCGGACCCACCCCTCTCCGGGCTGTACGGTGGCCGTGCAATACCTG 181
OY 33 1leAspAspAla1leAla1a1leGluGluGluSerGluThr1leAspProMet1yValPro 52
      ::::::::::::::::::::
DB 182 TTCGACACGAGTCTCAAGGCCATCACACGTCGGGGCGCGTGGACCGCTCACTGTGCCA 241
OY 53 Aspn1sAlaAsp1ySpheGluArgn1sVal1Gly1leVala1AspPhe1yGluLeuAla 72
      ::::::::::::::::::::
DB 242 AACACACGCGCGCTGACGACCAAGTTC---ATCGTGAAGGTGAGCCGACGGCGGGA 298
OY 73 MetArgen1leGluAlaArg1yLeu1yGluMet1yArgGluGlyAspAlaAsnVal 92
      ::::::::::::::::::::
DB 299 CTCACAAATGACGCGCTCACCGGCTGTCACCAATGCACAGCAGCGGTGAC----- 349
OY 93 LysGluGluGluGly1y1leVal1yAla1n1s1leu1leGly1yVal1HisAsp1leVal 112
      ::::::::::::::::::::
DB 350 -----TGCATGTGACTGTGGTGCCACGCG-----GTG 379
OY 113 SerMetGluTyrAspLeuAlaTyr1yLys1leuGlyAspLeu1n1sProThr----- 128
      ::::::::::::::::::::
DB 380 CAAATGCGCGTGGACCTAGGC-----GCCGCGCCGCTGCACGCGCAAGTATAGGGGTGAG 433
OY 129 -----Thrn1sVal1leSerAsp1leGlu1n1sPheVal 139
      ::::::::::::::::::::
DB 434 GTGAAGCTCAGCGCTTACACCGGACGTCAACAATATGTGTGATGCAACAATGACAG 493
OY 140 ValAla1eSer1eGlu1y1leSerAspGlu1yAsn1leThrMetThrSerPheGluVal 159
      ::::::::::::::::::::
DB 494 ATCATCGTCGAGCGCGTACAGAACCGCGCGCAAGCCGACCTGGAGGGTTCGCGTG 553
OY 160 ArgGluPheAlaAsnVala1yAla1n1s1leGly1yLeuSer1leLeuAspPro1lePhe 179
      ::::::::::::::::::::
DB 554 CACCGGCTGACGCGTGTGCGCGTGCACGCGTGCAGAGGCGTGGACCGCTGCGCGTCCG 613
OY 180 GlyVal1eSerAspVal1eThrAla1lePheGlu 191
      ::::::::::::::::::::
DB 614 AACACGCTCTCCGTGCTACTCACACAGTCTTGAA 649

```

RESULT 6 BO834649 683 bp mRNA linear EST 08-NOV-2002
BO834649
LOCUS BO834649
DEFINITION Po.ad_01G02_TEXF1 Psoroptes ovis mixed Psoroptes ovis cDNA clone
BO834649
ACCESSION BO834649.1 GI:22138963
VERSION
KEYWORDS EST
SOURCE sheep scab mite.
ORGANISM Psoroptes ovis
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Sarcopitidae; Psoroptidae;

REFERENCE 1 (bases 1 to 683)
Kenyon,F., Welsh,M., Parkinson,J., Whitton,C., Blaxter,M. and Knox
AUTHORS

TITLE
JOURNAL
COMMENT
Expressed sequence tag survey of gene expression in the scab mite
Psoroptes ovis - allergens, proteinases and free radical scavengers
Unpublished (2002)
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JF, UK.

Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared for Dr David Knox, Moredun Research
Institute, UK Sequencing was performed by Mark Welsh, ICAPB,
Edinburgh
PCR Primers
FORWARD: M13R
BACKWARD: T7PL
Plate: 01 row: G column: 02
High quality sequence stop: 541.
Location/Qualifiers
1. 683
/organism="Psoroptes ovis"
/db_xref="taxon:83912"
/clone="Po.ad_01G02"
/clone_11b="Psoroptes ovis mixed"
/sex="Mixed"
/dev_stage="Mixed (predominantly adult)"
/note="Vector: Lambda Triplex2; Site: 1: SfilA; Site: 2:
SfilB; Psoroptes ovis is the infective agent causing sheep
scab, a serious dermatitis in sheep. The cDNA was
obtained from mites harvested from the skin of previously
unexposed sheep harbouring a moderate challenge
infection. The cDNA was cloned using the Clontech Lambda
Triplex2 Smart cDNA cloning system."

BASE COUNT 217 a 100 c 138 g 228 t
ORIGIN

Alignment Scores:
Pred. No.: 6.23e-06 Length: 683
Score: 123.50 Matches: 48
Percent Similarity: 45.32% Conservative: 44
Best Local Similarity: 23.65% Mismatches: 80
Query Match: 11.56% Indels: 31
DB: 14 Gaps: 11

US-10-024-955-7 (1-213) x BO834649 (1-683)

```

OY 1 MetMet1ySphe---Leu1eul1eAla1aVala1aPheVala1aVala1aSerAlaAspPro 19
      ::::::::::::::::::::
DB 119 ATGATCAATATTTGTCGTATGTTTGTGTGTACACTGTCTGACATATGACCA 178
OY 20 1len1sTyrAsp1y1leThrGluGlu1y1leAsn1yAla1leAspAspAla1leAla 39
      ::::::::::::::::::::
DB 179 -----GATGACGTCAATTAATGGCCGATTCATATGGGCAT 217
OY 40 1leGluGluSerGluThr1leAspProMet1yValProAspn1sAlaAsp----- 56
      ::::::::::::::::::::
DB 218 TTATTCAAATCAAAAGAAATGTAATCCACTTAATTT-----CATCAAGATTATTAATTTA 271
OY 57 ---LysPheGluArgn1sVal1Gly1yLeu1yAspPhe1yGlyGluLeuAlaMetArgAsn 75
      ::::::::::::::::::::
DB 272 GAAGTGAAGAAATAATATGGACCAATGACCATATGATGATGATGATGATGATGATGATG 331
OY 76 1leGluAlaArg1yLeu1yGluMet1yAspArgGluGlyAspAlaAsnVal1yGlyGlu 95
      ::::::::::::::::::::
DB 332 GGTGAATTTGTTGGCCTTATATAGTTTAAGCCTATTTGGATGTTGAATGACCAATGAT 391
OY 96 GluGly-----1leVal1yAla1n1s1leu1leGly1yVal1HisAsp1leVal 112
      ::::::::::::::::::::
DB 392 GAATCGCGCAATATGTTGTGAAGTGCCTTAA-----ACACGTAAACAAATCATTA 445
OY 113 SerMetGluTyrAspLeuAlaTyr1yLys1leuGlyAspLeu1n1sPro-----ThrThrn1s 130
      ::::::::::::::::::::
DB 446 ---TTAGATGCTGATTTGATGCAAAATATGGTATACATACCAAGCTGGTGGTCCATTCG 502
OY 131 Val1leSerAsp1leGluAlaArg1yLeu1yVala1a1eAsn1yGlu1y1leSerAsp1yGlu 150
      ::::::::::::::::::::
DB 503 ATTTGTTGCCACATGAGATGAAATTTTATCTTGGTGTAAAGTGTGCA----- 556
OY 151 Asn1leThrMetThrSerPheGluVala1ArgGluPheAlaAsnVal-----ValAsn 167
      ::::::::::::::::::::

```

KEYWORDS EST.
SOURCE Amblyomma variegatum.
ORGANISM Amblyomma variegatum
Eukaryota: Metazoa: Arthropoda: Chelicerata: Arachnida: Acari:
Parasitiformes: Ixodida: Ixodidae: Amblyomma.
REFERENCE 1 (bases 1 to 716)
AUTHORS Nene,V., Lee,Y., Skilton,R., Quackenbush,J., Gardner,M. and Bishop
,R.
TITLE A partial gene index of ESTs from the salivary glands of Amblyomma
variegatum
JOURNAL Unpublished (2002)
COMMENT Contact: Vish Nene
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-610-5968
Fax: 301-838-0208
Email: nene@tigr.org
Seq primer: M13 reverse.
Location/Qualifiers
FEATURES
source 1..716
/organism="Amblyomma variegatum"
/db_xref="taxon:34610"
/clone="AVAA57"
/clone_1lb="AVSG"
/tissue_type="Salivary glands"
/dev_stage="Adult"
/lab_host="E.coli strain DH10B-Tona"
/note="Vector: pCMVSPORT6.0 ccdB; Salivary glands were
dissected on day five after initiation of feeding. Total
RNA was prepared using acid guanidium
thiocyanate-phenol-chloroform extraction. The cDNA library
was custom prepared by Invitrogen Corporation. Briefly,
first strand cDNA was primed using oligo(dT) containing a
NotI site. Size fractionated double stranded cDNA was
ligated to EcoRV-NotI cleaved vector and electroporated
into E.coli."

BASE COUNT 140 a 215 c 239 g 122 t
ORIGIN
Alignment Scores:
Pred. No.: 1.38e-08 Length: 716
Score: 143.50 Matches: 50
Percent Similarity: 38.25% Conservative: 33
Best Local Similarity: 23.04% Mismatches: 93
Query Match: 13.44% Indels: 41
DB: 13 Gaps: 7

US-10-024-955-7 (1-213) x BM290948 (1-716)
QY 5 LeuLeuIILaLaLaValaLaPhaValaIaValSerAlaAspProIleHis-----Tyr 22
Db 114 GTCGTGCTCTGCTGCTGCTGCTGCGACGACGAGCTTACCATGCGTG 173
QY 23 AspLysIleThrGluGlu-----IleAsnLysAla 32
Db 174 GAAGGGGTACGACGACGACCCCTCTCGGGCGCTGACGGTCCGTGTCATTCCTG 233
QY 33 IleAspAlaIleAlaIleGluGlnSerGluThrIleAspPromeTylsValPro 52
Db 234 TTCGACACGAGTCAACACCATCACACGCTGGGGCGCGCTGACCTGCTGCGCA 293
QY 53 ASPHSLAAspLysPheGluArgHisValGlyIleValAspPheLysGluLeuAla 72
Db 294 AACACCACCGCCACTTCGACCAAGTTC---ATGCTAAGGTGACGCGCGCGGA 350
QY 73 MetArgAsnIleGluAlaArgLysGlnMetLysArgGlnGlyAspAlaAsnVal 92
Db 351 CTCCTACAAATGAGACGCTCAGCGGGCTGCACACATGCACAGACGGGTGAC----- 401
QY 93 LysGlyGluGluGlyIleValLysAlaHisLeuLeuIleGlyAlaHisAspIleVal 112
Db 402 -----TGCATGTGACTGTGTGTCGCCAACGGG-----GTG 431

QY 113 SerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThr----- 128
Db 432 CAATTCGGCGGTGACTATGAC-----GCCGGCCCGGTGCACGACCATATGAGGGGTAG 485
QY 129 -----ThrHisValIleSerAspIleGlnAspPheVal 139
Db 486 GTGAAGCTCATGCGCTTCACCGCGGACGTGCACAAATATGTGTGATGTAACAACATGCAG 545
QY 140 ValAlaLeuSerLeuGluLeuSerAspGluLysAlaIleThrMetThrSerPheGluVal 159
Db 546 ATCATCGTCGACCGCTGACAGAACCCCGGCGACGCGGACGTGAGCGGTTCGCGTG 605
QY 160 ArgGlnPheAlaAsnValAlaAsnHisIleGlyLysSerIleLeuAspProIlePhe 179
Db 606 CACGCGCTCAGAGGTGTGCGCGGTGACGCTGCAAGGGCTGAGACCGGTGGCGTCCGCGCC 665
QY 180 GlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLys 196
Db 666 AACACGCTCTCCGTGCTACTCACACAGTCTGCAAGAGAGATCGACGC 716

RESULT 5
BM289894 649 bp mRNA linear EST 01-JUL-2002
LOCUS EST576428 AVSG Amblyomma variegatum cDNA clone AVAA74 5' end, mRNA
DEFINITION
sequence.
ACCESSION BM289894
VERSION BM289894.1 GI:21639861
KEYWORDS EST.
SOURCE Amblyomma variegatum.
ORGANISM Amblyomma variegatum
Eukaryota: Metazoa: Arthropoda: Chelicerata: Arachnida: Acari:
Parasitiformes: Ixodida: Ixodidae: Amblyomma.
REFERENCE 1 (bases 1 to 649)
AUTHORS Nene,V., Lee,Y., Skilton,R., Quackenbush,J., Gardner,M. and Bishop
,R.
TITLE A partial gene index of ESTs from the salivary glands of Amblyomma
variegatum
JOURNAL Unpublished (2002)
COMMENT Contact: Vish Nene
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-610-5968
Fax: 301-838-0208
Email: nene@tigr.org
Seq primer: M13 reverse.
Location/Qualifiers
FEATURES
source 1..649
/organism="Amblyomma variegatum"
/db_xref="taxon:34610"
/clone="AVAA74"
/clone_1lb="AVSG"
/tissue_type="Salivary glands"
/dev_stage="Adult"
/lab_host="E.coli strain DH10B-Tona"
/note="Vector: pCMVSPORT6.0 ccdB; Salivary glands were
dissected on day five after initiation of feeding. Total
RNA was prepared using acid guanidium
thiocyanate-phenol-chloroform extraction. The cDNA library
was custom prepared by Invitrogen Corporation. Briefly,
first strand cDNA was primed using oligo(dT) containing a
NotI site. Size fractionated double stranded cDNA was
ligated to EcoRV-NotI cleaved vector and electroporated
into E.coli."

BASE COUNT 124 a 195 c 214 g 116 t
ORIGIN
Alignment Scores:
Pred. No.: 3.56e-07 Length: 649
Score: 132.50 Matches: 49
Percent Similarity: 37.26% Conservative: 30
Best Local Similarity: 23.11% Mismatches: 92

AUTHORS Kenyon, F., Welsh, M., Parkinson, J., Whitton, C., Blaxter, M. and Knox, D.P.
TITLE Expressed sequence tag survey of gene expression in the scab mite *Psoroptes ovis* - allergens, proteinases and free radical scavengers
JOURNAL Unpublished (2002)
COMMENT Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 The library was prepared for Dr David Knox, Moredun Research Institute, UK Sequencing was performed by Mark Welsh, ICABP, Edinburgh

PCR Primers
 FORWARD: M13R
 BACKWARD: T7PL
 Plate: 06 row: C column: 07
 High quality sequence stop: 392.

FEATURES

source

1..601
 Location/Qualifiers
 /organism="Psoroptes ovis"
 /db_xref="taxon:83912"
 /clone_lib="Po_ad_06C07"
 /sex="Mixed"
 /dev_stage="Mixed (predominantly adult)"
 /note="Vector: Lambda Triplex2; Site_1: SfiI; Site_2: SfiI; Psoroptes ovis is the infective agent causing sheep scab, a serious dermatitis in sheep. The cDNA was obtained from mites harvested from the skin of previously unexposed sheep harbouring a moderate challenge infection. The cDNA was cloned using the Clontech Lambda Triplex2 Smart cDNA cloning system."

BASE COUNT 218 a 88 c 104 g 191 t
ORIGIN

Alignment Scores:

Pred. No.: 9.41e-16 **Length:** 601
Score: 196.00 **Matches:** 51
Percent Similarity: 45.21% **Conservative:** 34
Best Local Similarity: 27.13% **Mismatches:** 67
Query Match: 18.35% **Indels:** 36
DB: 14 **Gaps:** 5

US-10-024-955-7 (1-213) x B0835021 (1-601)

QY 2 MetLysPheLeuLeuIleAlaValAlaPheValAlaVal--SerAlaAspProIle 20
 |||||
 DB 84 ATGAATTTCTTCACATAC--TTGATCGGCTTATATGACATTTATTCAGGGCTTCTGT 140
 |||||
 QY 21 HsTYrAspLysIleThrGluIleAsnLysAlaIleAspAspAlaIleAlaIle 40
 |||||
 DB 141 CATGGGATGATAAACATGATCAAGTAAACCTTTCATACCAATACCTTCATCAATCAGTG 200
 |||||
 QY 41 GluInserGluThrIleAspProMetLysValProAspHisAlaAspPheGluArg 60
 |||||
 DB 201 TTAATAAATCGCACCGCTCGATTCATTAACCTTCACAAATCATATCAACAGATTGATCGC 260
 |||||
 QY 61 HsValGlyIleValAspPheLysGluLeuAlaMetArgAsnIleGluAlaArgGly 80
 |||||
 DB 261 AAAATTTGGTATCCCATATACATATTAATGATGCAAAATATTTGAATATATATGCT 320
 |||||
 QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGluGluGlyIleValLys 100
 |||||
 DB 321 TTAGGCGCATATACATCGTAAACGTAATGATCATCTTGAATTCGACAGCTGA----- 371
 |||||
 QY 101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
 |||||
 DB 372 -----CAATTCGATTTAAATTT 389

QY 121 LysLeuGlyAspLeuHis----- 126
 |||||
 DB 390 CATTAAATGATGATCATGATTCAGAAAGTATTCATATCAATTTAATATCAT 449
 |||||
 QY 127 -----ProTHrHisValIleSerAspIleGlnAspPheValAlaLeuSerLeu 144
 |||||
 DB 450 TCGTACCCGAAATTTGAAAGCTGAAATGATATCGGTGATATGCAATTAAGAAATTA 509
 |||||
 QY 145 GluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsn 164
 |||||
 DB 510 GCTGTCTTA---AGCGGTAATTTGACATTAATCATTTTGAAGTTGATGTAAGAAAC 566
 |||||
 QY 165 ValValAsnHisIleGlyLeu 172
 |||||
 DB 567 TGTAAAGTATATGATCATGATTTG 590

RESULT 2

B0835006 598 bp mRNA linear EST 08-AUG-2002
 LOCUS Po_ad_06B03.5
 DEFINITION Po_ad_06B03.5' similar to Q9UG12 PUTATIVE ALLERGEN PRECURSOR.
 Lepidoglyphus destructor, mRNA sequence.

ACCESSION B0835006.1 GI:22139320
 VERSION B0835006
 KEYWORDS EST.
 SOURCE sheep scab mite.
 ORGANISM Psoroptes ovis
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Acariformes; Sarcopitiformes; Astigmata; Sarcopitidae; Psoroptidae;
 Psoroptes.

REFERENCE

1 (bases 1 to 598)
 Kenyon, F., Welsh, M., Parkinson, J., Whitton, C., Blaxter, M. and Knox, D.P.

JOURNAL

Unpublished (2002)
 Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 The library was prepared for Dr David Knox, Moredun Research Institute, UK Sequencing was performed by Mark Welsh, ICABP, Edinburgh

COMMENT

Expressed sequence tag survey of gene expression in the scab mite *Psoroptes ovis* - allergens, proteinases and free radical scavengers
 Unpublished (2002)
 Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 The library was prepared for Dr David Knox, Moredun Research Institute, UK Sequencing was performed by Mark Welsh, ICABP, Edinburgh

FEATURES

source

1..598
 Location/Qualifiers
 /organism="Psoroptes ovis"
 /db_xref="taxon:83912"
 /clone_lib="Po_ad_06B03"
 /clone_lib="Po_ad_06B03"
 /sex="Mixed"
 /dev_stage="Mixed (predominantly adult)"
 /note="Vector: Lambda Triplex2; Site_1: SfiI; Site_2: SfiI; Psoroptes ovis is the infective agent causing sheep scab, a serious dermatitis in sheep. The cDNA was obtained from mites harvested from the skin of previously unexposed sheep harbouring a moderate challenge infection. The cDNA was cloned using the Clontech Lambda Triplex2 Smart cDNA cloning system."

BASE COUNT

214 a 86 c 107 g 191 t

ORIGIN

Alignment Scores: 8.13e-15 **Length:** 598
Pred. No.: 189.00 **Matches:** 51
Score:

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 14:55:03 : Search time 1475 Seconds
(without alignments)
2338.738 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 1068
Sequence: 1 MMKFLIAAFAVAVSADPI.....VRKEMTVLAPAFKRELEKN 213

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPRO.spool/US10024955/runat_14022003.130658.23717/app.query.fasta.1.391
-DB=EST -QEMT=fastap -SUFFIX=trst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10024955.ecgn.1.1.2874.0/runat.14022003.130658.23717 -MCPu=6 -ICPU=3
-NO.XLPHY -NO.MMAP -TARGET=NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :
EST : *
1: em_estba : *
2: em_esthum : *
3: em_estlin : *
4: em_estmu : *
5: em_estlov : *
6: em_estpl : *
7: em_estro : *
8: em_hc : *
9: gp_est1 : *
10: gp_est2 : *
11: gp_hc : *
12: gp_est3 : *
13: gp_est4 : *
14: gp_est5 : *
15: em_esthum : *
16: em_estom : *
17: gb_gss : *
18: em_gss_hum : *
19: em_gss_inv : *
20: em_gss_pln : *
21: em_gss_vit : *
22: em_gss_fun : *
23: em_gss_mam : *
24: em_gss_mus : *
25: em_gss_other : *
26: em_gss_pro : *
27: em_gss_rtd : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	196	18.4	601	B0835021	B0835021 Po.ad_06c
2	189	17.7	598	B0835006	B0835006 Po.ad_06b
3	170.5	16.0	516	B0834778	B0834778 Po.ad_03c
4	143.5	13.4	706	BM290948	BM290948 EST577482
5	132.5	12.4	649	B3289894	BM289894 EST576428
6	123.5	11.6	683	B0834649	B0834649 Po.ad_01c
7	121.5	11.4	604	B0834827	B0834827 Po.ad_03h
8	118.5	11.1	645	B0834612	B0834612 Po.ad_01b
9	116	10.9	614	B0835037	B0835037 Po.ad_06e
10	116	10.9	630	B0834844	B0834844 Po.ad_04b
11	103.5	9.7	673	B0038592	B0038592 DH01A03 H
12	100	9.4	537	TA86F08P	AL465038 T. brucei
13	100	9.4	617	AO654922	AO654922 Sheared D
14	96.5	9.0	625	AO651517	AO651517 Sheared D
15	96	9.0	523	BM276304	BM276304 PFEStoa8
16	96	9.0	572	BM275470	BM275470 PFEStoa8
17	93.5	8.8	533	TA175A02Q	AL475302 T. brucei
18	93	8.7	560	TA319G09Q	AL492823 T. brucei
19	92	8.6	952	B0644662	B0644662 AGENCOURT
20	91.5	8.6	944	B0943606	B0943606 AGENCOURT
21	91	8.5	673	B0583938	BE583938 6-7H-HA P
22	91	8.5	755	BH459040	BH459040 B06KD43TR
23	91	8.5	781	BG501777	BG501777 60254852
24	90.5	8.5	653	AI297539	AI297539 LP11861.5
25	90.5	8.5	722	BE536812	BE536812 601064811
26	90.5	8.5	754	B1104634	B1104634 602891509
27	89.5	8.4	920	AL514856	AL514856
28	89.5	8.4	929	BM461957	BM461957 AGENCOURT
29	89.5	8.4	982	B161010	B161010 601767637
30	88.5	8.3	507	TA2930782	TA2930782 474.dh258
31	88.5	8.3	603	BF668303	BF668303 602185166
32	88.5	8.3	623	BM044473	BM044473 603621970
33	88.5	8.3	649	AL134929	AL134929 DKE2P762B
34	88.5	8.3	678	BG748642	BE748642 602706344
35	88.5	8.3	688	BF136676	BF136676 601780441
36	88.5	8.3	688	B1917792	B1917792 603183892
37	88.5	8.3	707	BG775260	BG775260 602650274
38	88.5	8.3	802	BG367069	BG367069 HVSME1001
39	88.5	8.3	867	BQ923854	BQ923854 AGENCOURT
40	88.5	8.3	883	BQ686099	BQ686099 AGENCOURT
41	88.5	8.3	888	BE795760	BE795760 601590780
42	88.5	8.3	899	BQ940104	BQ940104 AGENCOURT
43	88.5	8.3	905	BQ063355	BQ063355 AGENCOURT
44	88.5	8.3	912	BQ431048	BQ431048 AGENCOURT
45	88.5	8.3	919	BQ689372	BQ689372 AGENCOURT

ALIGNMENTS

RESULT 1
B0835021
LOCUS B0835021
DEFINITION Po.ad_06c07 TEXT1 Psoroptes ovis mixed Psoroptes ovis cDNA clone
Po.ad_06c07.5, similar to Q9UGG2 PUTATIVE ALLERGEN PRECURSOR.
B0835021.1 GI:22139335
ACCESSION B0835021
VERSION B0835021
KEYWORDS Leptodoglyphus destructor, mRNA sequence.
SOURCE EST.
ORGANISM sheep scab mite.
Psoroptes ovis
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcoptiformes; Astigmata; Sarcopitoidae; Psoroptidae;
Psoroptes.
REFERENCE 1 (bases 1 to 601)

```
QY 91 -----AsnValIysGlyGluGlyIleValIleValIleHisLeuLeuIleGlyVal 107
Db 23746 TTACCTTTAAACGTTTCAGCTTATGTAATTAATTAATGCTTCA---ATTGAAGCATCA 23690
QY 108 HisAspAspIleValSerMetGluTyrAspIleAlaTyrIleLysLeuGlyAsp----- 124
Db 23689 CAT---CATTTCTATACCTTCTTGAATGATTTTAACTTCAAAGTTGGCAGCAAGAATTC 23633
QY 125 ---LeuHisProThrThrHisValIleSerAspIleGlnAspPheValVal----- 140
Db 23632 GAAGTCACACATTAATCATCTCAGACTTCACACATGTTGACTATGTGTTTAAGAAAA 23573
QY 141 -----AlaLeuSerLeuGluIleSerAsp----- 148
Db 23572 GGCACGATGATTACTAGCAATGAACAATTAACCATGATGATATGACATGCA 23513
QY 149 -----GluGlyAsnIle-----ThrMetThrSerPheGluVal 159
Db 23512 TACGTTAAAGAAATATTACAAACAAGAAATTCATTCACATCACTGACTTTCCTAAT 23453
QY 160 ArgGlnPheAlaAsnValVal----- 166
Db 23452 CCAGATTATGCGAACAGCCCATTAATAATTGTAACACGCTGGGATTTAGTGACGAAGA 23393
QY 167 ---AsnHisIleGlyLeuSerIleLeuAspProIlePheGlyValLeuSerAspVal 185
Db 23392 TGGAACCGTGCAGGTAAAGTAAGTGTGTA-----CGTGGCTGCTGA 23351
QY 186 LeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLysValLeuAlaProAla 205
Db 23350 ATGACTCCATACATGAACGAT-----AAAGAAGAAATTGATGCTTTACGCTCCTAAG 23300
QY 206 PheLysArgGluLeuGluLys 212
Db 23299 ATTCAAAGCTGCTTTACAGAGA 23279
```

Search completed: February 20, 2003, 18:30:58
Job time : 910 secs

```

      : LENGTH: 2718
      : TYPE: DNA
      : ORGANISM: Mus musculus
US-09-905-983-57

Alignment Scores:
Pred. No.:          5.07           Length:       2718
Score:             75.50           Matches:        47
Percent Similarity: 38.78%         Conservative:   48
Best Local Similarity: 19.18%      Mismatches:    85
Query Match:       7.07%           Indels:        65
DB:                10              Gaps:          9
US-10-024-955-7 (1-213) x US-09-905-983-57 (1-2718)

QY      18 AspProIleHis-----TyrAspLysIleThrGlu-----Glu 28
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      1078 GACCCCTCACAACATGGTCGCATTGACAAAGATGACCAGAAMAACAAGAGATCTGAGAGACAG 1137
QY      29 IleasnLysAlaIleaspsAlaIleAIA-----38
      ::::|:::|||||:::|||||:::|||||:::|||||:::
Db      1138 CTTCGGAAACCTGTGAATGATCAACACTCATCTTTCTTGAAACCAATGCCCTTG 1197
QY      39 -----AlaIleGluInserrGluThrIleaspprometLysValProAspHisAlasp 56
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      1198 CTGGTCTCTCTATTAGCGTGGCAAGACCGGGAATGAGAAAGAGTDAAGAAATACCCCAA 1257
QY      57 LysPheGluArgHisValGly--IleValAspPheLysGlyLysIleuAlaMetArgAsn 75
      |||:::|||||:::|||||:::|||||:::|||||:::
Db      1258 GTTTTCCGTCACATGCACAAACAGTGTGTGAGAGTTGCCAAATTTGGCTTTGCATCTCC 1317
QY      76 ILeGluAlaArgLysIleuLysGlnMetLys-----85
      |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1318 AACAAATGAGGAAGGGGTGAATTATGACAGAAATGCGACGCCACCAAGATTGACAGCTGTGT 1377
QY      86 -----ArgGln 87
      |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1378 CCCCAAGTCATTAACTGCTGCCCTCACACTGCTGCTCGGCCACAGATAAGTTCCTAG 1437
QY      88 GLyAspAlaAsnValLysGlyGluGluGlyIleValLysAlaHisLeuIleuIleGlyVal 107
      ::::|:::|||||:::|||||:::|||||:::|||||:::
Db      1438 GACAACATCGATGTCPTTCAAAAGACSGTAGGAAAAGCAAGTCCGTTGCTGTACTGAGCA 1497
QY      108 HisAspAspIleValSerMetGluTYrAspLeuAlaTYrLysLeuGlyAspLeuHisPro 127
      |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1498 GTGATGTACATCCACTCTGTGGATGTGACTCTCTCTGTCTCA-----1539
QY      128 ThrThrHisValIleSerAspIleGlnAspPheValAlaLeuSerLeuGluIleSer 147
      |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1540 GAAACCAATATCTTGGAGGATGTGGAACAAATGTGTGATGGC-----CTGCACAGAGGA 1593
QY      148 AspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValAlaAsn 167
      |||:::|||||:::|||||:::|||||:::|||||:::
Db      1594 GATGTGACACACTGGATGTCACACAGCTGGGGCCATACGGGGCGGCGGCAAGCGGTCTATT 1653
QY      168 HistIleGlyLeuSerIleLeuAspProIlePheGlyValLeuSerAsp--ValLeu 186
      |||||:::|||||:::|||||:::|||||:::|||||:::
Db      1654 CACATCTCATCATGACAGATGAGAGACTGTGAAGCGTGGGCTTAATACAGAGAAATGCTG 1713
QY      187 ThrAlaIlePheGlnAspThrValArgLysGluMetThrLysValLeuAlaProAlaPhe 206
      |||:::|||||:::|||||:::|||||:::|||||:::
Db      1714 GAACCC-----ACAAATATGCTTTCAGAAACAGTGTGCCACGCTTT 1755
QY      207 LysArgGluLeuGlu 211
      ::::::|||
Db      1756 GCTGACACAAGTTGAG 1770

RESULT 50
US-09-070-927A-138/c
; Sequence 138, Application US/09070927A
; Patient No. US2002012016M1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
```

```

1      Patrick J. Dillon
2      Steven Barash
3      TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
4      NUMBER OF SEQUENCES: 982
5      CORRESPONDENCE ADDRESS:
6      ADDRESSEE: Human Genome Sciences, Inc.
7      STREET: 9410 Key West Avenue
8      CITY: Rockville
9      STATE: Maryland
10     COUNTRY: USA
11     ZIP: 20850
12
13     COMPUTER READABLE FORM:
14     MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
15     COMPUTER: HP Vectra 486/33
16     OPERATING SYSTEM: MSDOS version 6.2
17     SOFTWARE: ASCII Text
18
19     CURRENT APPLICATION DATA:
20     APPLICATION NUMBER: US/09/070,927A
21     FILING DATE: 04-MAY-2000
22     CLASSIFICATION: <Unknown>
23
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER: 60/046,655
26     FILING DATE: 1997-05-16
27     APPLICATION NUMBER: 60/044,031
28     FILING DATE: 1997-05-06
29     APPLICATION NUMBER: 60/066,009
30     FILING DATE: 1997-11-14
31
32     ATTORNEY/AGENT INFORMATION:
33     NAME: Kenley K. Hoover
34     REGISTRATION NUMBER: 40,302
35     REFERENCE/DOCKET NUMBER: PB369
36     TELECOMMUNICATION INFORMATION:
37     TELEPHONE: (301) 309-8504
38     TELEFAX: (301) 309-8512
39
40     INFORMATION FOR SEQ ID NO: 138:
41
42     SEQUENCE CHARACTERISTICS:
43     LENGTH: 28690 base pairs
44     TYPE: nucleic acid
45     STRANDEDNESS: double
46     TOPOLOGY: linear
47
48     SEQUENCE DESCRIPTION: SEQ ID NO: 138:
49
50     US-09-070-927A-138
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

```

QY 78 AlaAag-----GlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnVal---LysGly 94
   |||:::|||||
Db 1183 GCCAAGGATGATCAGGTGTAAAGGCGCTGACGTGGCCAGCTTAAGCTTGTGACTCA 1242
QY 95 GlnGlnGlyIleValIleLysAlaHisLeuLeuIleGlyValHisAsp---AspIleValIser 113
   |||:::|||||
Db 1243 GAAGAAGATGCAAGTCAAGCCGCTTCTGACACATGAATAATCGTTCATGATGATGCTGCT 1302
QY 114 MetGlnTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrHisValIleSer 133
   |||:::|||||
Db 1303 GTTCGCTTT---GTTCGACCTTAAGGTGCTCTGTAATGCCCTGAG----- 1344
QY 134 AspIleGlnAspPheValAlaLeuSerLeuGlnIleSerAspGlnGlyAsn----- 151
   |||:::|||||
Db 1345 -----ATGCTGTACATTTCTTCAATGATGATGTTGTAAGAGTCAAGGAGAT 1389
QY 152 -----IleThrMetHisSerPheGlnValArgGlnPheAlaAsnValAlaAsn 167
   |||:::|||||
Db 1390 AAGGTGGCCCTCTGACGAGCGCCGTTTCTCTGTGTGTAATGCTGTGTGTGGA 1449
QY 168 HisIle-----GlyGlyLeuSerIleLeuAspProIlePheGlyVal 181
   |||:::|||||
Db 1450 CATATCGCTCTGAAGCTCAGATGTGTGA-----CCATTCGCTATCTC 1494
QY 182 LeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGlnMetThrLysVal 201
   |||:::|||||
Db 1495 CGTACCGCGGATGCTGTCAGCTTGACCAAGATACC-----AAGAATAATTCATGAGCC 1548
QY 202 LeuAlaProAlaPheLysArgGlnLeuGlnLys 212
Db 1549 GTATCC-----GAAGAAGAACTTGAAAAA 1572

```

RESULT 48 US-10-029-756-1

Sequence 1, Application US/10029756
Patent No. US20020108147A1

GENERAL INFORMATION:

APPLICANT: Thomas, Terry L.

TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
DELTA 6-DESATURASE

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Scully, Scott, Murphy & Presser

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: New York

COUNTRY: United States

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/029,756

FILING DATE: 21-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/934,254

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Presser, Leopold

REGISTRATION NUMBER: 19,827

REFERENCE/DOCKET NUMBER: 8383ZYXWVU

TELEPHONE: (516) 742-4343

TELEFAX: (516) 742-4366

TELEX: 230 901 SANS UR

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3588 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2002..3081
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-029-756-1

```

Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
	6.52	76.00	37.63%	23.66%	7.12%	3588	44	26	68	48	7

US-10-024-955-7 (1-213) x US-10-029-756-1 (1-3588)

```

QY 33 IleAspAspAlaIleAlaIleGlnIleSerGlnThrIleAspProMetLysValPro 52
   |||:::|||||
Db 743 CTGAATGATTCATCTCCTTGCGAGTCGCTTGTAGTGTGATGCGGCAAGTTACC 802
QY 53 AspHisAlaAspLysPheGlnArgHisValGlyIleValAspPheLysGlyLeuAla 72
   |||:::|||||
Db 803 GAT-----CGCATCATCATCATTCATTGTGGCGTGGGGAGTGAAC 844
QY 73 MetArgAsnIleGlu----- 77
   |||:::|||||
Db 845 ATGGCCATATTGAAAGAGTTAATTCACCGGCCATGAATTTGTGTATCGAAAGCAT 904
QY 78 -----AlaArgGlyLeu----- 81
   |||:::|||||
Db 905 ACAGATAATCGTTCTTGATGATGAGCGCCGCTGGGGGTGCCGTAATGTGGAGAT 964
QY 82 ---LysGlnMetLysArgGlnIleLysAlaAsnValLysGlyGlnGlyIleValLys 100
   |||:::|||||
Db 965 GCCCGCTGAGAAAGAGCTGTGGCTGGCCATATTCACCGAGCCGCAACCCATTGTG--- 1021
QY 101 AlaHisLeuLeuIleGlyValHisAspAspIleValIleSerMetGlnTyrAspLeuAlaTyr 120
   |||:::|||||
Db 1022 -----GTGGCCACCGACGACGACGCTTAACTTGAATTTGGCTTAAGTCC 1069
QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAspIleGlnAspPheValArg 140
   |||:::|||||
Db 1070 AAGGCG-----ATCGCCCGCTGACGAGTGTGGTTCGCTGGCAGAGATCCCAAGTT 1123
QY 141 AlaLeuSerLeuGlnIleSerAspGlnGlyAsnIleThrMetHisSerPheGlnValArg 160
   |||:::|||||
Db 1124 ACCCTGTCCTCCAGAGATATTGAATTGAACGGTGCCTTTCGCGCGGAATTTGCC 1183
QY 161 GlnPheAlaAsnValValaHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
   |||:::|||||
Db 1184 ACCATTCTCTTGGCGCGCGGCGGCTGGGGGC---AAATTTTGGGCACAC-----GCC 1234
QY 181 ValLeuSerAspValLeu 186
   |||:::|||||
Db 1235 ATGACCGATGATTTGCTG 1252

```

RESULT 49

US-09-905-983-57

Sequence 57, Application US/09905983

Patent No. US20020045591A1

GENERAL INFORMATION:

APPLICANT: Geiger, Benjamin

APPLICANT: Ben-Ze'ev, Avri

APPLICANT: Sadot, Elnat

TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER

FILE REFERENCE: 01/22326

CURRENT APPLICATION NUMBER: US/09/905,983

CURRENT FILING DATE: 2001-09-28

NUMBER OF SEQ ID NOS: 60

SOFTWARE: Patentin version 3.1

SEQ ID NO 57

Db	910	GAAGCAGCATT	921	111
Db	910	GAAGCAGCATT	921	111
RESULT	46			
US-09-974-300-476				
Sequence	476,	Application	US/09974300	
Patent No.	US20020146721A1			
GENERAL INFORMATION:				
APPLICANT:	Berka, Randy M.			
APPLICANT:	Clausen, Ib Groth			
TITLE OF INVENTION:	Methods For Monitoring Multiple Gene			
FILE REFERENCE:	10085.500-US			
CURRENT FILING DATE:	2001-10-05			
PRIOR FILING DATE:	2000-10-06			
PRIOR APPLICATION NUMBER:	60/279,526			
NUMBER OF SEQ. ID NOS:	8481			
SOFTWARE:	FASTSEQ for Windows Version 4.0			
SEQ ID NO	476			
LENGTH:	1287			
TYPE:	DNA			
ORGANISM:	Bacillus licheniformis			
US-09-974-300-476				
Alignment Scores:				
Pred. No.:	1.47	Length:	1287	
Score:	76.00	Matches:	48	
Percent Similarity:	42.39%	Conservative:	30	
Best Local Similarity:	26.09%	Mismatches:	58	
Query Match:	7.12%	Indels:	48	
Db:	10	Gaps:	10	
US-10-024-955-7 (1-213) x US-09-974-300-476 (1-1287)				
Oy	17	Alaasprrolienhstyaasplyllefthrcldnlu	-----lleasnlyala	32
Db	101	GGCGACCGCGTACTTTGAAACCGATTACACCCGAGATATTCATATGTCGTGAAGCG		160
Oy	33	-----lleasrpalaiallaa	-----	38
Db	161	GAAGCATCGACTTTGGCATTTGTGCATTCGCGGCCAATCCGCATCATGTATGAGCCAA	220	20
Oy	39	AlaalegluIn	-----SerGluthrlleasppromet	49
Db	221	GGCATTTGAAAGCGGCGCATCACCCCTGTTGGACATCATCGGAAACCTTTGACATCTT		280
Oy	50	LysvalrProasrphlsalaasplyrphedluarhlsvalgluyllevalasrphelysly		69
Db	261	GAAGACCGCGATCATTTTATACACTGCTTGACGAGCTCGGCATCTCCATGCCAAAGGA	340	40
Oy	70	Gluleualmetatargasnlleaglualatrgluleuysglmetlysatrglnllyasp		89
Db	341	GAAGCGCTTCATCAAAAGCGGAAAGCG	-----GCGAAAAA	376
Oy	90	Alaasnvallysglyuglnllylevalys	-----Alaahlsleuilellylvalnhs	108
Db	377	GGCAACCGAATCGGCTATCTGTGTATGACGGCGGTCAACGATCGAGCGGCGAATGCGC		436
Oy	109	Aspaarllevalisermetglutyaasrpeulatlutylusleuglyuasrleuhsrprothr		128
Db	437	ATGATGATCTTCACTCGAAAGCGGAGCTCGAAGAGCTCGTGGCGGGGCGCATGCCAATG		496
Oy	129	ThrhlvalllieseraspilleglasrphelyalalaleuSerleugluille	-----	146
Db	497	CGTTACCGGATTCG--ATCGATCAATACSTACGAGAAAGATTGAAATGAGACTT		55
Oy	147	---SeraspiluglyasnllethmetlthserphedgluvalatrglnrhealaaSnval		165
Db	554	GTCAACGATTCGCGGATCTTCAATCCGCACTAT	-----	588

```

QY      166 ValAsnHisIle 169
Db      590 ATCGAGCATATT 601

RESULT 47
US-09-815-242-9413
: Sequence 9413, Application US/09815242
: Patent No. US2002061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: EITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: fastseq for Windows Version 4.0
: SEQ ID NO 9413
: LENGTH: 1704
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1704)
US-09-815-242-9413

Alignment Scores:
Pred. No.: 2,21 Length: 1704
Score: 76.00 Matches: 58
Percent Similarity: 41.56% Conservative: 38
Best Local Similarity: 25.11% Mismatches: 81
Query Match: 7.12% Indels: 54
DB: 10 Gaps: 14

US-10-024-955-7 (1-213) x US-09-815-242-9413 (1-1704)
QY      1 MetcetylspheleuilellaValAlaPheValAlaValSerrilaasproIle 20
Db      985 GTTAAAGACTCTTGTTG-----GCAAAATGCTTC-----CTT 1017
QY      21 HlsTyraSplIsIleTr-----GlGluIleAsnLysAlaIleaspsAlaIle 37
Db      1018 CACGAGATGCCATCACTGATGCTACGTGTAAGACTGTGCTGAAACTGTGCTACTTTGCA 1077
QY      38 AlaAlaIleGluInserGluTrhIleasPProMetLysValProAspHisAlaasPls 57
Db      1078 GACTGACTCCAGGCCAAAGAAAGTTATCATGCGCATGGAAGAACCAAAAGCGGATGCT 1137
QY      58 PheGluATgHIsValIGlyIleValasPheLysGlyGluLeuAlaMetLrGAsnIleGlu 77
Db      1138 -----CCGCTTATCATCTTGAAACGGGAAACCTCTCCGACGTCAGTCT 1182

```



```

QY      109  --AspAspIleValSerMetGluTyrAspSerIleValTyrIleuGluValSerLeuIlePro 127
           |||::|||::  ::|||::|||::  |||::|||
Db      208  TTAAACCAAGCTTTACACAGCCSSCAACTGCTGTGTCATTCACAGCAAGCAAGATGATGACA 267
           |||::|||::  ::|||::|||::  |||::|||
QY      128  ThrThrIstValIleSerAspIle--GlnAspPheValAlaIaleuSerIleuIle 146
           ::|||::  ::|||::  |||::|||
Db      268  CCACGCTAAATTTATTCGACAGCAATTTGCTAAAGAT-----GCMAAGCATTTGAAATC 318
           ::|||::  ::|||::  |||::|||
QY      147  SerAsp-----GluGluAsnIleThrMetThrSerPheGluValAlaGluIlePheAla 163
           |||::|||::  ::|||::  |||::|||
Db      319  AAAGCTGGCTGATCGAAGCAAGAAAGTTTCTCACTGCA---CAATTACTGCGTTGGCA 375
           |||::|||::  ::|||::  |||::|||
QY      164  AsnValValAsnHisIleGlyGlyIleuSerIleuAspProIlePheGlyValIleuSer 183
           ::|||::  |||::|||::|||
Db      376  AAATCAACAACCCGCAAGGTTACTATCT-----GCMAAGCATTTGAAATC 405
           ::|||::  |||::|||::|||
QY      184  AspValIleuThrAlaIlePheGlnAspThrValAlaArgIleGluMetThrIleValIleAla 203
           ::|||::  |||::|||::|||
Db      406  --ATGCTATTATCTGATTATTAACAAGCCCACTCGCAAC-----GTGGCT 447
           ::|||::  |||::|||::|||
QY      204  ProAlaPheIysArgIleuGluIleAsn 213
           |||::|||
Db      448  TACGCTGTCAAGCAGCTGGCGCAGAAATAAAC 477
           |||::|||
RESULT 43
US-09-815-242-6517
; Sequence 6517, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6517
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(501)
; US-09-815-242-6517

```

Alignment Scores:	
Pred. No.:	0.322
Score:	76.50
Percent Similarity:	43.68%
Best Local Similarity:	26.84%
Length:	501
Matches:	51
Conservative:	32
Mismatches:	62

Query Match:	7.16%	Indels:	45
DB:	10	Gaps:	13
<hr/>			
US-10-024 -955-7 (1-213) x US-09-815-242-6517 (1-501)			
OY	38	AlAalalegluglnSerGluThrIleasprOmetLysValProAspHisAlaAspLys	57
		: ::	::
Db	10	GCAGCAATCGCTAAATAAGAAACCTTA-----GTTCAGCAGCCGCTGAGAA	57
OY	58	PheGu-----ArgHisValGlyIleValAspPheLysGlyLeuAlaMetLysasn	75
		:: ::	::
Db	58	TTTGATACGACGACATCGTGTATTCTGGACATCCCTGGT-----TTAACACTT	108
OY	76	IleGuaIlaTrGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyLu	95
		:: ::	
Db	109	GAAGAAGTCACTAACTTAGCTTAACAATTAGCT-----GATGCCGGCTA-----GAA	156
OY	96	GlUGlyIleValLysAlaHisLeu-----IleGlyValHis--	108
		: :: ::	:
Db	157	ATGAAGATCATCAAAAACCTAACTCTTCTCGTGCAGCTAAAAAGTTGGTTAGACGC	216
OY	109	--AspaspIleValSerMetGluTryTrspLeuAlaTryLysLeuGlyAspLeuHisPro	127
		:: :: :: :: ::	::
Db	217	TTAGACGAAGCTTATACAGGCCCAACTCTGTCGATTCAGCAACGACGAGATGATACCA	276
OY	128	ThrThrHisValIleSerAspIle---GlnAspPheValAlaIleSerLeuGluIle	146
		::	::
Db	277	CCAGCTAAATATTATCGACGAATTTCTCTTAAGAT-----GCCAAAGCATTTAGAAATC	327
OY	147	SerAsp-----GluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAla	163
		:: ::	
Db	328	AAAGGTGGCGTCATCGAAGAAAGATTCTTCACAGGAA--CAATATTACGCTGGCA	384
OY	164	AsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGlyValLeuSer	183
		:: ::	:: ::
Db	385	AAACTACCAACCCGCGAAGTTTACTACT-----	414
OY	184	AspValLeuThrAlaIlePheGlnAspPhrValArgLysGlnMetThrLysValLeuAla	203
		:: :: :: ::	:: ::
Db	415	---ANGCTATTATCTGTATTACAAAGCCGACGCGCACAC-----CTGGCT	456
OY	204	ProAlaPheLysArgGluLeuGlnLysasn	213
Db	457	TACGCTGTCAAAGCAGTGGCAGAGAAACAAC	486
<hr/>			
RESULT 44			
US-09-350-756-5/C			
Sequence 5, Application US/09350756			
Patent NO. US20020034521A1			
GENERAL INFORMATION:			
APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases			
APPLICANT: John S. Lee			
APPLICANT: Peter Pushko			
APPLICANT: Michael D. Parker			
APPLICANT: Jonathan F. Smith			
APPLICANT: Mark T. Dertzbaugh			
APPLICANT: Leonard Smith			
TITLE OR INVENTION: Botulinum Neurotoxin Vaccine			
FILE REFERENCE: 003/124/SAP RIID 96-21			
CURRENT APPLICATION NUMBER: US/09350,756			
EARLIER FILING DATE: 1999-07-09			
EARLIER APPLICATION NUMBER: US 60/092,416			
EARLIER FILING DATE: 1998-07-10			
NUMBER OF SEQ ID NOS: 11			
SOFTWARE: Apple Macintosh Microsoft Word 6.0			
SEQ ID NO 5			
LENGTH: 1317			
TYPE: DNA			
ORGANISM: Artificial			
FEATURE:			
OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"			
US-09-350-756-5			


```
US-10-024-955-7 (1-213) x US-09-738-626-1873 (1-735)
QY 5 LeuLeu1leAla1aVala1aPheVala1aValSer-----AlaAspPro 19
   |||||
Db 46 GTCTTAATTTCCACACTGACTTTGCCCTTCGAGCATAGAGGTGCGATCCA 105
QY 20 IleHisTyr--AspLysIleThr-----GluGluIleAsnLysAla--- 32
   |||||
Db 106 ATACATATTACCGATCAATTAATTAATGATTACACGCGCTGACAAATGTAAGTTGCT 165
QY 33 -----IleAspSpsAla1leAla1le 40
   |||||
Db 166 GATGCCCTTCGTCATACAACTCCCTGAAAGACCAAGTTAATTAAGTAAGAAACATC 225
QY 41 GluGlnSerGlu-----ThrIleAspProMetLysValProAspHisAlaAspLys 57
   |||||
Db 226 GAACAGAGGTGACACTGCGGATTCATCGACCCACAAACATCCG----- 270
QY 58 PheGluArgHisValGlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGlu 77
   |||||
Db 271 -----GTGGACATTTGAGGACTTTGAAATGAAACAGGAGGAAGGTACTATTGAG 318
QY 78 -----AlaArgGly-----LeuLysGlnMetLysArgGlnGlyAspAlaAsn--- 91
   |||||
Db 319 CGTTTGTGATGCGTACTGCTGATGTTCTACACTTCACATCCCGATTCGCAAAATAAAT 378
QY 92 -----ValLysGlyGlu-GluGlyIleValAlaHisLe 103
   |||||
Db 379 TCTCTGATCCAGCGGAGGTGATGTTCCCTGCGCTCATTTGGGATTTGAGGAATTCCTTT 438
QY 103 uLeuIleGlyValHisAspSpsIleValSerMetGluTyrAspLeuAlaTyrLysLeuG 123
   |||||
Db 439 GCTCAAGGCGATCGTTTAACTAATGCTTGGCACAAACACTATTATTATTTGGCTA-- 496
QY 123 yAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValValAlaLeuSe 143
   |||||
Db 497 -----CGATGTCAT-----TTCGTTTGGCT----- 517
QY 143 rIleuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAl 163
   |||||
Db 518 -----ACTATTACACACTATACAGATTCGGAATTCAG 549
QY 163 aAsn-----ValValAsnHisIleGly--GlyLeuSerIleLeuAspProIlePheGly 180
   |||||
Db 550 AACTTTTACATCCGTCATCACTTACTGGGATCGAATGAGTTTGAACACCAACATTTGAG 609
QY 181 ValLeuSerAsp-----ValLeuThrAlaIlePheGln 191
   |||||
Db 610 CACACTAGTGAATGAAATGTACGCATGAGTGTCTTTGTCATTTATTATTCAG 663

RESULT 40
; US-09-815-242-6903
; Sequence 6903, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6903
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2169)
; US-09-815-242-6903

Alignment Scores:
Pred. No.: 2,32 Length: 2169
Score: 77.00 Matches: 43
Percent Similarity: 42.86% Conservative: 41
Best Local Similarity: 21.94% Mismatches: 86
Query Match: 7.21% Indels: 26
DB: Gaps: 9

US-10-024-955-7 (1-213) x US-09-815-242-6903 (1-2169)
QY 7 IleAla1aVala1aPheVala1aValSerAlaAspProIleHisTyrAspLysIleThr 26
   |||||
Db 1576 ATTCGCGAGCATGTCGCGAGTATCCATTAATGATGAAGCAACCCATCGGTGATTCGACGTGACG 1635
QY 27 GluGluIleAsnLysAlaIleAspAspAlaIleAlaIleGlu-----GlnSerGlu 44
   |||||
Db 1636 GATACCTTTAA--AACGATAGCTTCACGCGCATTCACGCTTACCAAGCA 1686
QY 45 ThrIleAspProMetLysValProAspHis-----AlaAspLysPheGluArg 60
   |||||
Db 1687 AATATTGATGTGCTAATTATGAGTGGCGATCAACAATCTCTCGGTTCATGACGGA 1746
QY 61 HisValGlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGly 80
   |||||
Db 1747 CACTCGGCATTTAAGAAAGCTTTGGGAACCTAATCCAGAGAT---AAAGGGAACAA 1803
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyIleValLys 100
   |||||
Db 1804 ATCCAAAGCTGAAGATCTAGTCAATATTGTGCAATGCTAGCGCATGGAATAAATGAT 1863
QY 101 AlaHisLeuLeuIleGlyValHisAspAspIle---ValSerMetGluTyrAspLeuAla 119
   |||||
Db 1864 GCGCGCTTACCTTACCTTTCGAAACGTCAGTTTGGCATGAATCAAGTTCAGATTTCGA 1923
QY 120 TyrLysLeuGlyAspLeuHisProThrHisValIleSerAspIleGlnAsp----- 137
   |||||
Db 1924 GAACAAACAGCTCTGCAACGCTTATGCAACACTTCGGTGAATCACTTGGATGCGCTT 1983
QY 138 PheValVala1aLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPhe 157
   |||||
Db 1984 TTTATTGCAAGACCAACTTGAA--AATATCAACAAATCTAATT 2028
QY 158 GluValArgGlnPheAlaAsnValVala1aHisIleGly-----GlyLeuSer 173
   |||||
Db 2029 -----TTTGTCTTGATTTACAAATATTCGCGCTACCCCTTGGCGCTTTCGCT 2076
QY 174 IleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIle 189
   |||||
Db 2077 TTTCTTAAGCCCTAATCGCTGCTGCTCATGCGGTGAGTCATTT 2124

RESULT 41
; US-09-870-759-105
; Sequence 105, Application US/09870759
```

Db 351 GCACCTGCCACTCTCGACACGAGCGCCAGGGGCACTCCGCCCTCGGAGACTCTCC 292
Qy 92 Vallysglylu-----Gluglyllevallysalahisleuile 105
||| |||||
Db 291 GTGAGAGGGGAGATCGTACTCGGTGACACATGCGGGCCGCCCGGTAGTTGTGACCTC 232
Qy 106 GlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrIlyLeuGlyAspLeu 125
||| |||||
Db 231 GCGCTCATGAGGAGTACCGCGGTACAGACAGTACTCGCGGGC-----GGGCAAGCG 178
Qy 126 HisProThrThrHisValIleSerAspIleGlnAspPheValAlaLeuSerLeuGlu 145
Db 177 CACGCGCGCGGGTACTTGTGATGACGGCGAA-----CTCGGG 139
Qy 146 IleSerAspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnVal 165
||| |||||
Db 138 TCCGAGACCTTGTGCGATCGCCCTCTGC-----GACTCAGAGATTCTTC---ATA 88
Qy 166 ValAsnHisIleGly 170
||| |||||
Db 87 GTTGAACACCTCGGT 73

RESULT 38

US-10-124-800-27/C
; Sequence 27, Application US/10124800
; Publication No. US20020194641A1
; GENERAL INFORMATION:
; APPLICANT: Metz, James
; APPLICANT: Barclay, William
; APPLICANT: Platt, James
; APPLICANT: Kumer, Jerry
; TITLE OF INVENTION: PUFA Polyketide Synthase Systems and Uses Thereof
; FILE REFERENCE: 2997-29
; CURRENT APPLICATION NUMBER: US/10/124,800
; CURRENT FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 09/231,899
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/284,066
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/298,796
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/323,269
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 4512
; TYPE: DNA
; ORGANISM: Schizochytrium sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4512)
; OTHER INFORMATION:
US-10-124-800-27

Alignment Scores:

Pred. No.: 5.78 Length: 4512
Score: 77.50 Matches: 42
Percent Similarity: 38.388 Conservative: 29
Best Local Similarity: 22.70% Mismatches: 69
Query Match: 7.26% Indels: 45
DB: 9 Gaps: 8

US-10-024-955-7 (1-213) x US-10-124-800-27 (1-4512)

Qy 9 AlaValAlaPheValAlaValSerAlaAspProIleHisTyrAspIlySerGluGlu 28
||| |||||
Db 561 GCGCGCGTTCAGCTACGACGCTGCTGAGAGAACATGAGATGCCCGCTCGAGAGC 502
Qy 29 IleAsnIlyAlaIleAspAspAlaIleAlaAlaIleGluGlnSerGluThrIleAspPro 48
||| |||||
Db 501 CTTGGCGAAGCCGCTGACGCGGAATGTC-----GTACTCGAGGGT 463

Qy 49 MetIlyValProAspHisAlaAspIlyPheGluArg----- 60
||| |||||
Db 462 CTCGCCCTGCTGGGCGACCGCGTAAAGGTGCTGTGAGCAGCGGTAGACGGC 403
Qy 61 -----HisValGlyllevallysalahisleuile 71
||| |||||
Db 402 GTGCCCTGTTCTGGAAGTCAATGCCCATGTAGGA-----GATGAGCATGAGATC 352
Qy 72 AlaMetArgAsnIleGluAlaArgGlyLeuIlyGlnMetIlyArgGlnIlyAspAlaAsn 91
||| |||||
Db 351 GCACCTGCCACTCTCGACACGAGCGGCCCGGAGGCACTCCGCCCTCGGAGAGCTCTC 292
Qy 92 Vallysglylu-----Gluglyllevallysalahisleuile 105
||| |||||
Db 291 GTTGAAGGGGAGATCGTACTCGGTGACACATGCGGGCCGCCCGGTAGTTGTGACCTC 232
Qy 106 GlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrIlyLeuGlyAspLeu 125
||| |||||
Db 231 GCGCTCATGAGGAGTACCGCGGTACAGACAGTACTCGCGGGC-----GGGCAAGCG 178
Qy 126 HisProThrThrHisValIleSerAspIleGlnAspPheValAlaLeuSerLeuGlu 145
Db 177 CACGCGCGCGGGTACTTGTGATGACGGCGAA-----CTCGGG 139
Qy 146 IleSerAspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnVal 165
||| |||||
Db 138 TCCGAGACCTTGTGCGATCGCCCTCTGC-----GACTCAGAGATTCTTC---ATA 88
Qy 166 ValAsnHisIleGly 170
||| |||||
Db 87 GTTGAACACCTCGGT 73

RESULT 39

US-09-738-626-1873
; Sequence 1873, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1873
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1873

Alignment Scores:

Pred. No.: 0.483 Length: 735
Score: 77.00 Matches: 60
Percent Similarity: 40.00% Conservative: 36
Best Local Similarity: 25.00% Mismatches: 59
Query Match: 7.21% Indels: 85
DB: 9 Gaps: 14

```

; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3687
; LENGTH: 3089
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X07173
US-09-880-107-3687

```

```

Alignment Scores:
Pred. No.: 3.33 Length: 3089
Score: 77.50 Matches: 57
Percent Similarity: 35.76% Conservative: 46
Best Local Similarity: 19.79% Mismatches: 84
Query Match: 7.26% Indels: 101
DB: 10 Gaps: 14

```

US-10-024-955-7 (1-213) x US-09-880-107-3687 (1-3089)

```

QY 19 PROLEHISTYRAspLysIleThrGluGluIleAsnLysAlaIleAsp-----AspAla 36
   ||| :||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 964 CCTGACACCTGGACCCCAATTCGCCAAACATCCTTTGTCTATCATCATGTGAGTGGCTGC 1023
QY 37 ILeAlaAlaIleGluGlnSerGluThrIleAspPromeLysValProAspHisAlaAsp 56
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1024 ATGTGGGAGTAAATGAACAACTGTGAAGCAATGAAAGCATATTGTGATGACCTC 1083
QY 57 LysPheGluArgHisValGlyIleValAspPheLysGlyGluLeu---AlaMetArgAsn 75
   :||| :||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1084 AAGAGCAAGACCAATTCTGTGTGATTGATTCACCAAGACCACTTGCAACTGGAGAAAT 1143
QY 76 -----IleGluAlaArgLysLeuLysGlnMet 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1144 GATTTATTTCAGCTACAAAACACACAGTTGCAGATGACCAAGAGGTATATTGGAGAAATC 1203
QY 85 LysArgGlnGlyAspAlaAsnValLysGlyGluGlyIleValLysAlaHisLeuLeu 104
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1204 CAGCCCAAGTGAAGGACCAACAATC-----AACGAAAGCACTCCTACGGCAATCTTCATT 1257
QY 105 ILeGlyValHisAspAspIleValSerMetGlu-----TyrAspLeuIleTyrLys 121
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1258 TTGAATGAAGCCCAATTAAGTGGACTGTTAGACCCCACTCCGTCTCGTGATCATTTTG 1317
QY 122 LeuGlyAspLeuHisProThrThr-----HisVal 131
   :||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1318 GTTTTCGATGGAGATCCAAACAGTGGGCGAAGTAACCTGTCAAAATTCAGAAAACGTT 1377
QY 132 ILeSerAspIleGlnAspPheValValAlaLeuSerLeu----- 144
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1378 AAGGAGACATCCACACATATATCTCTGTTCAAGTTTGGCGATGGCATTTGATGTGGAC 1437
QY 145 -----GluLeuSerAspGlu-----Gly 150
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1438 TATGATTTTTCAGACAGCTGTCCAAATGAAGAACCATGCAATTCACAAAGGATTTTGG 1497
QY 151 AsnIleThrMetHisSerPheGluValArgGlnPheAlaAsnValVal----- 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1498 AAC---CAGGACACGCTCTCCAGCTTAAGAAATTTCTACACACAGGTCTCCATCATG 1554
QY 166 ----- 166
DB 1555 CTCGGGAATGTCAGTTCACTATCCCATACATCAGTCACGAGCGTCACTCAAAACAT 1614
QY 167 -----AsnHisIleGlyGlyLeuSerIleLeu-----AspProIle--- 178
   ||| :||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1615 TTCATTAATCACTGTGGAGGCTCAGAGATTGTGTGGCAGGAAATTTTGACCTGCTAAA 1674

```

```

QY 179 PheGlyValLeuSerAspValLeuThrAlaIle----- 189
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1675 TTGGATCAAAATGAGAGCGTTATATACGCGACTTGCGGTATACACGAGTATAGCTTGAG 1734
QY 190 -----PheGlnAspThrValArgLysGluMetThrLysValLeu 202
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1735 ACCCTGGCCAGATGAGACACTTGCAGGATTTTCTATCGAAAGAC-----AAGCATGCA 1788
QY 203 AlaProAlaPheLysArgLysLeu 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1789 GATCCCGATTTCACCAAGAACTG 1812

```

RESULT 37

```

US-10-124-800-5/c
; Sequence 5, Application US/10124800
; Publication No. US20020194641A1
; GENERAL INFORMATION:
; APPLICANT: Metz, James
; APPLICANT: Barclay, William
; APPLICANT: Flact, James
; APPLICANT: Kumer, Jerry
; TITLE OF INVENTION: PURA Polyketide Synthase Systems and Uses Thereof
; FILE REFERENCE: 2997-29
; CURRENT APPLICATION NUMBER: US/10/124,800
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 09/231,899
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/284,066
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/298,796
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/323,269
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 4509
; TYPE: DNA
; ORGANISM: Schizochytrium sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4509)
; OTHER INFORMATION:
US-10-124-800-5

```

```

Alignment Scores:
Pred. No.: 5.77 Length: 4509
Score: 77.50 Matches: 42
Percent Similarity: 38.38% Conservative: 29
Best Local Similarity: 22.70% Mismatches: 69
Query Match: 7.26% Indels: 45
DB: 9 Gaps: 8

```

US-10-024-955-7 (1-213) x US-10-124-800-5 (1-4509)

```

QY 9 AlaValAlaPheValAlaValSerAlaAspProIleHisTyrAspLysIleThrGluGln 28
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 561 GCGCGCGTGGAGCTAGACAGTACTCGAAGAAACATGAGATCCGCGCTGAGAG 502
QY 29 ILeAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspPro 48
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 501 CTTGGGAGAGCCGCGTACCGCAATGTC-----GTACTCGAGGCT 463
QY 49 MetLysValProAspHisAlaAspLysPheGluArg----- 60
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 462 CTCGCCCTCGTGGCCACGCCGCTAAAGCTGAGCGTGTGTGAGCAGCGCGTACGCGG 403
QY 61 -----HisValGlyIleValAspPheLysGlyGluLeu 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 402 CTCGCCCTGTTCTGGAAGTCAATGCCCATGTAGGA-----GATAGCATGAGATC 352
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 72 AlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsn 91
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

: PRIOR APPLICATION NUMBER: 96/09339
: PRIOR FILING DATE: 1996-07-19
: PRIOR APPLICATION NUMBER: PCT/FR97/01326
: PRIOR FILING DATE: 1997-07-16
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 13
: LENGTH: 1362
: TYPE: DNA
: ORGANISM: Infectious bursal disease virus
US-09-784-990-13

Alignment Scores:
Pred. No.: 0.874 Length: 1362
Score: 78.00 Matches: 39
Percent Similarity: 45.24% Conservative: 37
Best Local Similarity: 23.21% Mismatches: 64
Query Match: 7.30% Indels: 28
DB: Gaps: 10

US-10-024-955-7 (1-213) x US-09-784-990-13 (1-1362)
Oy 40 IleglInserGluThrIleAspPrometLysValProAspHisAlaSprLysPheGlu 59
::: ::::||| :::::
Db 304 GTGACTCGGAGTCTCAAGTACAGTCACACACACACTCCCTGGCGGCTTATGCACTAAAC 363
Oy 60 ArgHisValGlyIleValAspPheLysGlyGluLeuAla--MetArgAsnIleGluAla 78
::: ::||| :::::|
Db 364 GGCACCATTAACCGCGGTGACCTCCAGAGACCGTGAAGTGAAGTCAAGATGTAGCTAAC 423
Oy 79 ArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyIle 98
||| |||
Db 424 AATGCGGTG-----ATGCTGCACACGCCAACATCAACGACAAATTTGG-- 468
Oy 99 ValLysAlaHisLeuLeuIleGlyValHisAspArgIleValSerMet-----GluTyr 116
469 -----AACTCTCTGTATGGGAGAGGGCTCACTGTCTCTAGCTACCCACATCATAT 519
Oy 117 AsPLeuAlaTyr--LysLeuGlyAspLeuHisProThrThrHisValIleSeraspIle 135
||||| ||| ::::|
Db 520 GATCTGGGTATGTACAGGCTTGTCACCCCATTCGCCGTATPAGGCTT----- 567
Oy 136 GlnAspPheValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIle--ThrMet 154
||| ::| ::|
Db 568 --GACCCAAAATGTGTAGCTACATGCAGACAGCATGACAGGCCACAGATCTACACATA 624
Oy 155 ThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyIleLeuSerIle 174
625 ACTGCAGCGGATGTACCAATCTCATCACTACAGTCAACCA--GGTGGGTACCAATC 681
Oy 175 LeuAspProIlePheGlyValLeuSerAspValLeuThrAla----- 188
682 -----ACACTGTTTCAACCCAACTTGTATGCTATCAACAGCTCAGCATTTGGGGAGAG 735
Oy 189 ---IlePheGlnAspThrValArg 195
:::||||| ::::|
Db 736 CTCGTGTTTCAAAACAAGGCTCCAA 759

RESULT 35
US-09-960-428-13
: Sequence 13, Application US/09960428
: Patent No. US20020115147A1
: GENERAL INFORMATION:
: APPLICANT: Roche Diagnostics GmbH
: TITLE OF INVENTION: Method for producing an active heterodimeric AMV-RT in prokaryotic
: FILE REFERENCE: 5272/00/
: CURRENT APPLICATION NUMBER: US/09/960.428
: CURRENT FILING DATE: 2001-09-21
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 13
: LENGTH: 2155
: TYPE: DNA

```

[illegible]

```
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(648)
US-09-815-242-8986

Alignment Scores:
Pred. No.: 0.297      Length: 648
Score: 78.00      Matches: 52
Percent Similarity: 38.91%      Conservative: 41
Best Local Similarity: 21.76%      Mismatches: 90
Query Match: 7.30%      Indels: 36
DB: 10      Gaps: 10

US-10-024-955-7 (1-213) x US-09-815-242-8986 (1-648)

QY 1 MetMetLysPheLeuLeuLeuAlaValAlaPheValAlaValSerAlaSprProIle 20
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATGAAATAAGATGATATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 60

QY 21 -----HistyrAspLysIleThrGluLeuIleAsnLysAlaIleAspAla 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GCTGGCTTTATTTTGGATTAACACACACACTATGTA-----ACAACGTATATGCA 114

QY 37 lIeAlaAlaIleGluGlnSerGluThrIleAspPro----- 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 AAGATAAATGGGATCAATTAATAAGCAAGTCCGATCGATCGACAAATTAATCATT 174

QY 49 ---MetLysValProAspHisAlaAspLysPheGluArgHisValGlyIleValAspPhe 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 AATGTTAAACAAGGTGACGAACTCGATTAAGGTGACAAA---GTACACATTTGTTACTGTA 231

QY 68 LysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGln 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 CAGAGTCAAGATGCG-----GAACGAAAGATATGATCTTAATAATGCTCAAAAAGGT 285

QY 88 GlAspAlaAsnValLysGlyGluGly---IleValLysAlaHisLeuLeuIleGly 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 ACAATTCGCAAAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 345

QY 107 ValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHis 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 -----TATGCATCAATTTAGATGATTTATAT 372

QY 127 ProThrThrHisVal---IleSerAspIleGlnAspPheValValAlaLeuSerLeuGlu 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 GTAACGACAAATATGACGAAAGATATTAAGATGTTGAAGTGAAGATGTTGAT 432

QY 146 lIeSerAspGlu-----GlyAsnIle 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 GTGACATATGACGGTCAAAAAGCTTCTATTAAAGTAAAGTTGATGATGTAAGAAAGA 492

QY 153 ThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeu 172
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 ACTGTCGCAAGCTTTTCATATGATGATGATGATGATGATGATGATGATGATGATGAT 552

QY 173 SerIleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 TCTCAAGTAATACCTGTT-----AAAATC 576

QY 193 ThrValArgLysGluMetThrLysValLeuAlaProAlaPheLysArgLysLeuGlu 211
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 ACTCTAAGATCAAGAACCGTCTTAACAAGTTGTTCTCGAATGAACCGCTGAAGTAA 633

RESULT 33
; US-09-938-842A-812
; Sequence 812, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
```

```
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 812
; LENGTH: 768
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-812

Alignment Scores:
Pred. No.: 0.381      Length: 768
Score: 78.00      Matches: 27
Percent Similarity: 46.98%      Conservative: 43
Best Local Similarity: 18.12%      Mismatches: 63
Query Match: 7.30%      Indels: 16
DB: 9      Gaps: 3

US-10-024-955-7 (1-213) x US-09-938-842A-812 (1-768)

QY 7 lIeAlaAlaValAlaPheValAlaValSerAlaAspProIleHisTyrAspLysIleThr 26
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 GTGAAGCGGTGAGCTGTGGATGGGACATGATGATGATGATGATGATGATGATGATGAT 259

QY 27 GluGluIleAsnLysAlaIleAspAlaIleAlaAlaIleGluGlnSerGluThrIle 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 GACATGCTGTGCTGTGCGGAGGAGGAGGTGCTGCGCTGAGGAGGAGGAGGAGGAGGAG 319

QY 47 AspProMetLysValProAspHisAlaAspLysPheGluArgHisValGlyIleValAsp 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 GAAGCGGTAAAGGTGTGAGCTGCGATGACGAGCGGAGGAGGAGGAGGAGGAGGAGGAG 379

QY 67 PheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMet----- 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 GTGGAG---GAGGCGGACAGCGCGCGGTGAGAGGTGCTGTGCTGTGCTGTGCTGTGCTG 436

QY 85 -----LysArgGlnGlyAspAlaAsnValLysGly 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 437 GATACGCTGTGACACAGGTGCTGTGCTGTGACGAGAGATATGAGAGTGGAGGTGCCGGG 496

QY 95 GluGluGlyIleValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMet 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 GACATGAGAGTGTGTGAGCGGTGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 556

QY 115 Glu-----TyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisVal 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 GAGGAGCTGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 616

QY 132 lIeSerAspIleGlnAspPheValVal 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 GTGCTGAGCTGTGTGACATGTGTGTG 643

RESULT 34
; US-09-784-990-13
; Sequence 13, Application US/09784990
; Patent No. US20020037292A1
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, JEAN-CHRISTOPHE
; APPLICANT: BOUCHARDON, ANNABELLE
; APPLICANT: RIVIERE, MITCHEL
; TITLE OF INVENTION: AVIAN POLYNUCLEOTIDE VACCINE FORMULA
; FILE REFERENCE: 454313-2260
; CURRENT APPLICATION NUMBER: US/09/784,990
; CURRENT FILING DATE: 2001-02-16
```


US-09-738-626-1074

Alignment Scores:
Pred. No.: 3.11 Length: 3627
Score: 78.50 Matches: 39
Percent Similarity: 41.14% Conservative: 33
Best Local Similarity: 22.29% Mismatches: 68
Query Match: 7.35% Indels: 35
DB: 9 Gaps: 6

US-10-024-955-7 (1-213) x US-09-738-626-1074 (1-3627)

```
OY 13 ValAlaValSerAlaAspProIleHisTryAspLysIleThrgluGluIleAsnLysAla 32
   ::::::::::::::::::::
DB 133 TTGGTGGTACTGCTACTGCGCGTAGGCTGAGATCTCCAGCGGTGAGTTGAAGCGGATG 192
OY 33 IleAspAspAlaIleAlaAlaIleGluGlnSerGluThrIleAspProMetLysValPro 52
   ::::::::::::::::::::
DB 193 ATGGGCGATMAAGTGGCGGCTCCGCTCGTGGAGACTGCGCACATGAGCGATTGAGT 252
OY 53 AspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyIleuAla 72
   ::::::::::::::::::::
DB 253 CCGGCTGTGACATTTGGGCAAGCGTGTCTCAGTGTCTCAATCATCTAGCGCACTCAAA 312
OY 73 MetArgAsnIleGluAlaArgGlyLeuLysGln-----MetLysArgGlnGlyAsp 89
   ::::::::::::::::::::
DB 313 GTGCGTGGTGGTGGTGGCGCGCTGTTTCCAGCGGTACTCAAGATGCTGAAGTGTCT 372
OY 90 AlaAsnValLysGlyGluGlnGly-----
   ::::::::::::::::::::
DB 373 GCACCCCTCCTCTTAAAGAGCGGAGTTTGAATTTCTGCCCTGACCACTGAGCTG 432
OY 98 IleValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTryAsp 117
   ::::::::::::::::::::
DB 433 GTTTTCCGCGCTAC-----AAGCATGTCTGATGTGTGCGCAAGCGCGTGAAG 480
OY 118 LeuAlaIleLysLeuGly-----AspLeuHisProThr-----
   ::::::::::::::::::::
DB 481 TTGGTACCCGCGGTGGCATTTTGGATATTTCCACACATTTGGATTATCCGGTGTCT 540
OY 130 -----HisValIleSerAspIleGlnAspPheValAlaIleuSerLeu 144
   ::::::::::::::::::::
DB 541 GTGAGATTCTGGGCGATGAAGTCTGTGACATTAGGCACTGTCTGTCCCGATCAGCC 600
OY 145 GluIleSerAspGluGlyAsnIleThrmethIleThrmethPheGluVal 159
   ::::::::::::::::::::
DB 601 ACCATTCCAGAA-----ATCACCATCAAGACGATTGAGATT 636
```

RESULT 30

US-09-070-927A-17/c

Sequence 17, Application US/09070927A

Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ. ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 32768 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 17:

US-09-070-927A-17

Alignment Scores:
Pred. No.: 75.8 Length: 32768
Score: 78.50 Matches: 52
Percent Similarity: 37.50% Conservative: 38
Best Local Similarity: 21.67% Mismatches: 85
Query Match: 7.35% Indels: 65
DB: 10 Gaps: 11

US-10-024-955-7 (1-213) x US-09-070-927A-17 (1-32768)

```
OY 5 LeuLeuIleAlaIleAlaValAlaPheValAlaValSerAlaAspProIleHisTryAspLys 24
   ::::::::::::::::::::
DB 11469 TTAGTGATGGATGAAGAGCGCCATCTTATATACCTTTCATCCAGCTATGTTTAA 11410
OY 25 IleThnGluGluIleAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGlu 44
   ::::::::::::::::::::
DB 11409 GAAGATCTTGAGTTTAATCAATG-----AACCTATGCTTTCAGTTGAACAA 11362
OY 45 ThrIleAspProMetLysValProAspHisAlaAspLysPheGluArgHisValGly--- 63
   ::::::::::::::::::::
DB 11361 -----CCAAATAAATTAAAGATGCTGATATTTTGAAGAAACAAACAAACGAT 11311
OY 64 -----IleValAspPheLysGlyIleuAlaMetArgAsnIleGlu 77
   ::::::::::::::::::::
DB 11310 CAAATTGATGCTTTTATATCGCTGACTAT----- 11281
OY 78 AlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGly 97
   ::::::::::::::::::::
DB 11280 -----TTTCGCATCCCAAGGCAAGTCAATCATATTAAAGAGCAATATTC 11233
OY 98 IleValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTryAsp 117
   ::::::::::::::::::::
DB 11232 CTTCGCCCTCCACACTTA-----ACTAGAACGCGTTATATCAG 11197
OY 118 LeuAlaIleLysLeuGlyAspLeuHisProThrHisValIleSerAspIleGlnAsp 137
   ::::::::::::::::::::
DB 11196 CTAAATTAACACGCTG-----GTTCGTACCAAGCAACTTATGAAAACATTAATAT 11143
OY 138 PheValAlaIleLeuSerLeuGluIleSerAspGluGlyAsnIleThrmethIleSer--- 156
   ::::::::::::::::::::
DB 11142 AATGCAACACGCTTTCTTAAGACACTAAAGCAGAGGTGATCCGTTTAACTGCCACA 11083
OY 157 -----PheGluValArg 160
   ::::::::::::::::::::
DB 11082 TTAGTAACCTAATGACGGAAGATTACACATGAGCAACACTTCCGGAAGCTGATTAAGA 11023
OY 161 GlnPheAlaAsnValAlaAsnHisIleGlyGlyLeuSerIleLeuAspPro-----Ile 178
   ::::::::::::::::::::
DB 11022 GAGTTTGCAGAACCTAATTCAAAACACTAGGAGAGTATCGTTTAAAAATCCCGAAGAAAT 10963
```


Db	1825	AAA-----CCATATCTCGTAATTTATTTGGATGAAATGAAAAACACAT	1869
Qy	137	AspHeValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetHis	156
		:::	:::
Db	1870	CCAGATGATTTATTAATTCATTACCAAGTTTGATGACGACATTGGACATCAAAA	1929
Qy	157	PheGluValArgInPheAlaAsnValVal-----AsnHisIleGlyGlyLeu	172
			:::
Db	1930	GGAGCTACAGTTGATTTCCACAAATACACATTATCATTAATGACATCAACGTTGACACAA	1983
Qy	173	SerIleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGln	192
		:::	:::
Db	1990	GAAATTACAGATCCAAACATTTGCTGGATTCGGTGGCTTCAAGTACGACACAAATATGAA	2049
Qy	193	ThrValAlaGlyGluMetThrIlyValLeuAlaProAlaPheIlyArgIle	209
			:::
Db	2050	ACAATTGCAAAAACGATGTTAAAGAAATTAAAAATTCATTCCTGCTCAAA	2100

RESULT 28
US-09-815-242-8691
; Sequence 8691, Application US/09815242
; Patent No. US20020061569A1
; Current: 11/20/2002

```

1: GENERAL INFORMATION:
2: APPLICANT: Haselbeck, Robert
3: APPLICANT: Ohlson, Kari L.
4: APPLICANT: Zyskind, Judith W.
5: APPLICANT: Wall, Daniel
6: APPLICANT: Trawick, John D.
7: APPLICANT: Carr, Grant J.
8: APPLICANT: Yamamoto, Robert T.
9: APPLICANT: Xu, H. Howard
10: TITLE OF INVENTION: Identification of Essential Genes in
11: TITLE OF INVENTION: Prokaryotes
12: FILE REFERENCE: ELITRA.011A
13: CURRENT APPLICATION NUMBER: US/09/815.242
14: CURRENT FILING DATE: 2001-03-21
15: PRIOR APPLICATION NUMBER: 60/191,078
16: PRIOR FILING DATE: 2000-03-21
17: PRIOR APPLICATION NUMBER: 60/206,848
18: PRIOR FILING DATE: 2000-05-23
19: PRIOR APPLICATION NUMBER: 60/207,727
20: PRIOR FILING DATE: 2000-05-26
21: PRIOR APPLICATION NUMBER: 60/242,578
22: PRIOR FILING DATE: 2000-10-23
23: PRIOR APPLICATION NUMBER: 60/253,625
24: PRIOR FILING DATE: 2000-11-27
25: PRIOR APPLICATION NUMBER: 60/257,931
26: PRIOR FILING DATE: 2000-12-22
27: PRIOR APPLICATION NUMBER: 60/269,308
28: PRIOR FILING DATE: 2001-02-16
29: NUMBER OF SEQ ID NOS: 14110
30: SOFTWARE: FastSeq for Windows Version 4.0
31: SEQ ID NO 8691
32: LENGTH: 2457
33: TYPE: DNA
34: ORGANISM: Staphylococcus aureus
35: FEATURE:
36: NAME/KEY: CDS
37: LOCATION: (1)..(2457)
38: 09-09-815-242-8691

```

Alignment Scores:	
Pred. No.:	1.77
Score:	78.50
Percent Similarity:	39.63
Best Local Similarity:	22.58
Query Match:	7.35
DB:	10
Length:	245
Matches:	49
Conservative:	37
Mismatches:	86
Indels:	45
Gaps:	11

US-10-024-955-7 (1-213) x US-09-815-242-8691 (1-2457)

```

QY      18 AspproilehstYrAspLysileThrglu-----IleasnLysala 32
      ||| ::::||| ||| ::::: |||:::|||||

```

Db	1510	GATCATTACATGAGAGATTATTGGGCAAAAAAGATGCTGTTAAATTCATACAGTAAACG	1569
Qy	33	ILeaspaspalailealaiailegluGlnSerGIuThIleasprometLysValPro	52
		::: ::::	
Db	1570	GTTTGACGCTCCCGTGAGGGTTAAAA	1602
		::::: ::::	-----GATCCT-----
Qy	53	AspHisAlaAspLysPheGluArgHisValGIleValAspPheLysGIuIleuAla	72
		::: ::::	
Db	1603	-----AAACGACCAATTGGTAGCTTTATCTCTCTTGAGACCAACTGCT	1644
Qy	73	MetArgasnIleGlu--AlaArgGIuLeu--LysGlnMetLysArgGlnLysAla	90
		:::	
Db	1645	GTTGCTAAACTGATATAGCTAGACGCTTTAGCGATCAATGTTGGCGATGATGATCGC	1704
Qy	91	AsnValLysGIuLeu--GIuGIuIleValLysAlaHisLeu-----IeuIleLeu	106
		::: :::: ::::	:
Db	1705	ATGATCCGCTGACATGACGACTGAATTTATGGAAAAACGACGATGACCGGATAGTGGT	1764

Qy 107 Val-----Hisaspapillevelsernetcluryspneulatyr 120
 |||||
Dd 1765 GCTCCTCCAGCATATGTTGGTCATGATGATGGTGCACAAATTACTGAAAAAGTTAGACGT 1822

Oy	121	LysLeuGlyAspLeuHisProThrHisValIleSerAsp-----LleGln	136
Db	1825	AAA-----CCATATCTGTAAATTATTTTGATGAATAATGAAAAACCTCAT	1869
Oy	137	AspPheValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrmethrSer	156
Db	1870	CCAGCATGTATTTAATATCTATTACAAGTTTTAGATGATGCACATTTGCACATCAAAA	1929
Oy	157	PheGluValArgGlnPheAlaAsnValVal-----AsnHisIleGlyGlyLeu	172
Db	1930	GGACGTACACTGTCATTTCAGAAATTACAAATTATTCATATGACATCACAAGCTTGCGCACAA	1989
Oy	173	SerIleLeuaspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp	192
Db	1990	GAAATTACAAATCACAAGATTTCGTGGATTCGGTGGTTCACATGATGAGCAACAAGATTATGAA	2049
Oy	193	ThrAlaIrrLySGluMetThrLysValIleuAlaProAlaPheLysArgGlu	209
Db	2050	ACAAATTCGAAAACGATGTTAAAGAATTAATAAATTCTTCCTCGTCCGAA	2100

RESULT 29
US-09-738-626-1074
; Sequence 1074, Application US/09738626
; Publication No. US20020197605A1

1 GENERAL INFORMATION:
 2 APPLICANT: NAKAGAWA, SATOSHI
 3 APPLICANT: MINOGUCHI, HIROSHI
 4 APPLICANT: ANDO, SEIKO
 5 APPLICANT: HAYASHI, MIKIRO
 6 APPLICANT: OKOI, KEIKO
 7 APPLICANT: YOKOI, HARUHIKO
 8 APPLICANT: TATEISHI, NAKO
 9 APPLICANT: SENOH, AKIHIRO
 10 APPLICANT: IKEDA, MASATO
 11 APPLICANT: OZAKI, AKIO
 12 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 13 FILE REFERENCE: 249-125
 14 CURRENT APPLICATION NUMBER: US/09/738, 626

1 PRIOR APPLICATION NUMBER: JP 99/377484
 2 PRIOR FILING DATE: 1999-12-16
 3 PRIOR APPLICATION NUMBER: JP 00/159162
 4 PRIOR FILING DATE: 2000-04-07
 5 PRIOR APPLICATION NUMBER: JP 00/280988
 6 PRIOR FILING DATE: 2000-08-03
 7 NUMBER OF SEQ ID NOS: 7059

; SEQ ID NO 1074

```

;
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
;

```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 4531
; LENGTH: 2430
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4531

Alignment Scores:
Pred. No.: 1.74 Length: 2430
Score: 78.50 Matches: 49
Percent Similarity: 39.63% Conservative: 37
Best Local Similarity: 22.58% Mismatches: 86
Query Match: 7.35% Indels: 45
DB: 10 Gaps: 11

US-10-024-955-7 (1-213) x US-09-815-242-4531 (1-2430)
QY 18 Aspproilehistrasplysilethrglu-----lleasnlysa 32
DB 1510 GATACATTACAGAGAGATTGTGGCAAAAGATGCTGTAATTCATCAGTAAGCG 1569
QY 33 lleasppalallealaleleuglnsergluthrllaspprometlyvalpro 52
DB 1570 GTTACAGTCCCGCGGACGGTTAA-----GATCCT----- 1602
QY 53 Asphsialasplysphegluarghisvalglylevalasphelysglyleuala 72
DB 1603 -----AAACGCCAATTGCTAGCTTATCTCTTGACCAACTGCT 1644
QY 73 Metargasnllleglu---Alaarglyleu---Lysglmetlyarglnclyaspala 90
DB 1645 GTTGCTAAACGTAATGCTAGCTTACCTGATCAATCAATCGCGCATGATGCG 1704
QY 91 Asnvallysglylu---gluglylevallysalaHisleu-----leuilegly 106
DB 1705 ATGATCCGCTGATGACATGATGTAATTTATGGAACACGCGATGACCGATGCTG 1764
QY 107 Val-----HisaspspillevalsermetclutyraspleuAlaTyr 120
DB 1765 GCTCCTCCAGATATGTTGGTATGATGATGCGATCAATCAAGCAAAAGTTAGACT 1824
QY 121 LysleuglyaspheuhisprothrrhithisvalIleSerASP-----llealn 136
DB 1825 AAA-----CCATATTCCTGTAATTTATTTATGATGAATGCAAAAGCACAT 1869
QY 137 AsphevalvalaleuSerleuIleSeraspGluIleSeraspGluIleSerasp 156
DB 1870 CCAGATGATTTAATATCTTATACAGTTTATGATGACGACATTTGACATCAAAA 1929
QY 157 PhegluvalarglnphealasnvalVal-----AsnHisileglyglyleu 172
DB 1930 GGACGTACAGTTGATTTCAAGAAATACATATATCAATGACATCAAGCTTGACACA 1989
QY 173 SerIleuaspProilephegllyvalleuSeraspvalleuThrAlaIlePhegl 192
DB 1990 GAATTTACAGATCAAGATTTGCTGATTCGTTGCTGATCAAGTATGACACATTA 2049
QY 193 ThrvalarglysglumethrlyvalleuAlaProAlaPheIlyargly 209
DB 2050 ACAATTCGAAAAACGATGTTAAAGAAATTAATAAAATTCATCCGCCAGAA 2100

RESULT 27
US-09-815-242-8326
; Sequence 8326, Application US/09815242
; Patent No. US20020061369A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

```

```

; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 8326
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2457)
US-09-815-242-8326

Alignment Scores:
Pred. No.: 1.77 Length: 2457
Score: 78.50 Matches: 49
Percent Similarity: 39.63% Conservative: 37
Best Local Similarity: 22.58% Mismatches: 86
Query Match: 7.35% Indels: 45
DB: 10 Gaps: 11

US-10-024-955-7 (1-213) x US-09-815-242-8326 (1-2457)
QY 18 Aspproilehistrasplysilethrglu-----lleasnlysa 32
DB 1510 GATACATTACAGAGAGATTGTGGCAAAAGATGCTGTAATTCATCAGTAAGCG 1569
QY 33 lleasppalallealaleleuglnsergluthrllaspprometlyvalpro 52
DB 1570 GTTACAGTCCCGCGGACGGTTAA-----GATCCT----- 1602
QY 53 Asphsialasplysphegluarghisvalglylevalasphelysglyleuala 72
DB 1603 -----AAACGCCAATTGCTAGCTTATCTCTTGACCAACTGCT 1644
QY 73 Metargasnllleglu---Alaarglyleu---Lysglmetlyarglnclyaspala 90
DB 1645 GTTGCTAAACGTAATGCTAGCTTACCTGATCAATCAATCGCGCATGATGCG 1704
QY 91 Asnvallysglylu---gluglylevallysalaHisleu-----leuilegly 106
DB 1705 ATGATCCGCTGATGACATGATGTAATTTATGGAACACGCGATGACCGATGCTG 1764
QY 107 Val-----HisaspspillevalsermetclutyraspleuAlaTyr 120
DB 1765 GCTCCTCCAGATATGTTGGTATGATGATGCGATCAATCAAGCAAAAGTTAGACT 1824
QY 121 LysleuglyaspheuhisprothrrhithisvalIleSerASP-----llealn 136

```



```
DB: 10 Gaps: 8
US-10-024-955-7 (1-213) x US-09-815-242-6804 (1-2157)
QY 5 LeuLeuIIleAlaValAlaPheValAlaValSerAlaAspProIleHIStryAspLys 24
   ||| ||| :||| |||
Db 1033 TTAGCAATAGGTGGCTTACGCTTTGCAATGCGGGTGTAAAGACCTTGAAGTTCCTGAT 1092
QY 25 ILeHrGIuIIleAsnLysAlaIleAspAlaIleAlaIleIleGIuInserGIu 44
   :||| ||| ||| ||| :||| :|||
Db 1093 GTTACAAATGAAACGAAACGGCCGCTTACAAAGCCCTTCAAAAGTCTGGCTGAAAGTC 1152
QY 45 ThrIleAspPrometLysValProAspHisAlaAspLysPheGIuArgHisValGIuIle 64
   :||| :||| ||| ||| |||
Db 1153 GATAGTAAACCAAAAATTCGCCGAC-----GATTAAGATTGAA----- 1191
QY 65 ValAspPheLysGIuIleuAlaMetArgAsnIleGIuAlaArgGIuLysGIuMet 84
   :||| :||| :||| :||| :|||
Db 1192 -----GAAAGCAAGGTGTCAAAACAGACCCCGCAAAATCATCTGTGAAAAA 1242
QY 85 LysArg-----GIuGIuAspAla 90
   |||
Db 1243 GACCGATCTGTTACTTTGTACATCACTCTGACACAGAAAAAATTGAGATGCCGATTAT 1302
QY 91 AsnValLysGIuGIuGIuGIuIleValLysAlaHisLysLeuIleGIuValHisAspAsp 110
   :||| :||| :||| :||| :|||
Db 1303 ACAACGAAATCGATGATCTGCTGTGCAAGCCTTCAAAAACCTAGCGTTTTCAGAAAGC 1362
QY 111 IleValSerMetGIuIlyrAspLeuAlaIlyrLysLysGIuAspLeuHisProThrHis 130
   :||| :||| :||| :||| :|||
Db 1363 CAATTTTACACGAAAAAAGATATGACT-----GATTCTGTGTCTACAGATATAC 1410
QY 131 ValIleSer-----AspIleGIuAspPheValAla 141
   :||| :||| :||| :||| :|||
Db 1411 ATTATTAAACAAAAACCACTGACGTAAAAAAAGTTGATCCGAAAAAAGCAAGTCACT 1470
QY 142 LeuSerLeuGIuIleSerAspGIuLysAsnIleHrMetHrSerPheGIuValArgGIu 161
   :||| :||| :||| :||| :|||
Db 1471 TTAAACGCTAGTGAACGACAGCAGAACGCG---GTTACTTGTGCTTGTATACGCGGTTATTC 1527
QY 162 PheAlaAsnValAlaHisHisIleGIuGIuLysLeuSerIleLeuAspProIlePheGIuVal 181
   :||| :||| :||| :||| :|||
Db 1528 TACACAAATGACAGTAATGACTGCGACAGTTA-----GGCAT 1566
QY 182 LeuSerAspValLeuThrAlaIlePheGIu-----AspThrValAlaArgLysGIuMetThr 199
   :||| :||| :||| :||| :|||
Db 1567 TCTGAATCTCAATTTACGCGGTGTGCGACCAAGCAGATACGATACAGCAGTTTATGTC 1626
QY 200 LysValLeuAlaProAla 205
   ||| |||
Db 1627 ATTACGCAAGACCCGCCA 1644

RESULT 22
US-10-147-026-15
; Sequence 15, Application US/10147026
; Publication No. US20030003538A1
; GENERAL INFORMATION:
; APPLICANT: Dietrich, Paul Shartzer
; APPLICANT: Huang, Chiao-Chiao
; APPLICANT: Johnson, Carl D.
; APPLICANT: Sangameswaran, Lakshmi
; TITLE OF INVENTION: Neuropathic Pain Genes; Compositions
; FILE REFERENCE: ROCH-006
; CURRENT APPLICATION NUMBER: US/10/147,026
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/155,702
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/189,931
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 15
```

```
; LENGTH: 18946
; TYPE: DNA
; ORGANISM: human
US-10-147-026-15

Alignment Scores:
Pred. No.: 25.3 Length: 18946
Score: 79.50 Matches: 41
Percent Similarity: 37.24% Conservative: 32
Best Local Similarity: 20.92% Mismatches: 70
Query Match: 7.44% Indels: 53
DB: 9 Gaps: 8

US-10-024-955-7 (1-213) x US-10-147-026-15 (1-18946)
QY 26 ThrGIuGIuIIleAsnLysAlaIleAspAlaIleAlaIleIleGIuInserGIuThr 45
   ||| ||| :||| |||
Db 8062 ACTGTTAGAGTGAACATTTTGGCCAAATGTGCGAGAAATTGTGACTTTGACAGCA 8121
QY 46 IleAspPrometLysValProAspHisAlaAspLysPheGIuArgHisVal----- 62
   :||| :||| :||| :||| :|||
Db 8122 GCTTCAGATCTGTCTATAGTATGATGAAGAGAAATTTTCAATTCATTCATGATAGAACT 8181
QY 63 -----GIyIleValAspPheLysGIuGIuLeuAlaMetArgAsnIleGIuAla 78
   :||| :||| :||| :||| :|||
Db 8182 TTCCCTGTGTCAGAGAAATGTTACTGTTAACCTGAAATTTATGGCAAAATCTAGAACTC 8241
QY 79 ArgGIuLysLysGIuMetLysArgGIuGIuLysAspAlaAsnValLysGIu----- 95
   :||| :||| :||| :||| :|||
Db 8242 -----AATTTTGTAACTTTAGCGGCAAACTTTTCTTT 8274
QY 96 ---GIuGIuIleValLysAlaHisLysLeuIleGIuValHisAspIleValSerMet 114
   :||| :||| :||| :||| :|||
Db 8275 CCTGAGGGGCGTGTGAATACAAACATTTGTTGTCATTTGTGGATGACAAACATTCCTGAG 8334
QY 115 GIu-----TyrAspLeuAlaIlyrLysLysGIuGIuLysLeuGIuAspLeuHis 126
   :||| :||| :||| :||| :|||
Db 8335 GAGAAAGAGTATACCAACTCATCTGTATGATGTC-----AGGACACAGAGATTCCA 8388
QY 127 ProThrThrHisValIleSerAspIleGIuAspPheValAlaIleLeuSerLeuGIuIle 146
   :||| :||| :||| :||| :|||
Db 8389 CCAGCCGGAATCCCTGCTGTGATGCTCAAGATATGACAGTCTCTCACAGTAAAGCC 8448
QY 147 SerAsp----- 148
   ||| |||
Db 8449 AGTGATGAACCAATGAGATTAAATTTGCTTTCATCAAGATTGTGTTACTACAA 8508
QY 149 GIuGIuAsnIleHrMetHrSerPheGIuValArgGIuPheAlaAsnValAlaHis 168
   :||| :||| :||| :||| :|||
Db 8509 GAGGCTAATACATAATTCAGCTTTTCATCAACAGAGAAATTTGATCT----- 8556
QY 169 IleGIuGIuLysLeuSerIle-----LeuAspProIlePheGIuValLeuSer 183
   :||| :||| :||| :||| :|||
Db 8557 CTAGAGCTATCAATGTCAATATACCAAGGTTCTGTGAATGCTGAGT 8604

RESULT 23
US-09-738-626-2343
; Sequence 2343, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
```

```

Db 1470 GGTCTGCTGGCGGACGAGGGGTGGAGAGGTCCACAGCGCGGAGCTGACCGAGTCTC 1411
Qy 97 -----glyIleValIysAlaHisLeuLeuIleGlyVal 107
Db 1410 GAGAACGTCTTCCTCGCTGGCGGACTCGGCCCTCGTT---GCCGAGCTTTCCTGGTGGC 1354
Qy 108 HisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisPro 127
Db 1353 CACCTCGACCTCGTCGCGCGGACGAGCTGAGCTTGGCC-----GGCTCGCTGACCA 1303
Qy 128 ThrThrHisValIleSerAspIleGlnAspIheValAlaLeuSerLeuGluIleSer 147
Db 1302 GCGCCGAGTCCGTCGCGCCGCCACCAAGCTGGCGGTGATGCACAGCTCCTGATGTCGCG 1243
Qy 148 AspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValAlaSn 167
Db 1242 CATCCAGGAGAGATCATCTTCTGCTACGCTTGGGCGACGAAGTGGATACGGCCGCTTTC 1183
Qy 168 HisIleGlyIleuSerIleLeuAspProIlePheGly 180
Db 1182 CAC---GGCGGC-----GATGGCGTCTTCGCG 1159

```

RESULT 20

```

US-09-738-626-1968
; Sequence 1968, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1968
; LENGTH: 2493
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1968

```

Alignment Scores:

```

Pred. No.: 1-15 Length: 2493
Score: 80.00 Matches: 43
Percent Similarity: 43.42% Conservative: 23
Best Local Similarity: 28.29% Mismatches: 53
Query Match: 7.49% Indels: 33
DB: 9 Gaps: 8

```

US-10-024-955-7 (1-213) x US-09-738-626-1968 (1-2493)

```

Qy 29 ILeAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspPro 48
Db 289 ATCAACCTGCTGTCACAGCATATTCCTGCTGATCCAGCAAGACTTCTTGACAGC 348
Qy 49 -----MetLysValProAspHisAlaAspLysPheGluArg----- 60
Db 349 GATGTCGTGTCGTCATTCCTGAT---GCAGATCTTTCGATCGCATGATGATGATTGAT 405

```

```

Qy 61 -----HisValGlyIleValAspPheLys-----Gly 69
Db 406 TTACTGCTGTAACAAGACACGCTGACGAGTGGCTGCAATTGCACTGACACGGGT 465
Qy 70 GLeuLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAsp 89
Db 466 GAGCTGTCGTGCTTTAAGTGGCCGCGA-----AATAGC 501
Qy 90 AlaAsnLysGlyGluGluGlyIleValAlaHisLeuLeuIleGlyValHisAsp 109
Db 502 GCTTGGCACGCGGTCACATTTTTCGAGGTCGCGGATGTCCTTAGCGGTGCTGTCAC 561
Qy 110 -----AspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisPro 127
Db 562 CTCGAAGACGTGGTTCATTTTGAGGTC---ATTGCCACTTCTTAGAGGTGATGAG 618
Qy 128 ThrThrHisValIleSerAspIleGlnAspIheValAlaLeuSerLeuGlu---Ile 146
Db 619 ACCTGAGAGTGCATATTCATGACGAAATGGCGAGTGCCTTATAGTCTTGAGGCTGTC 678
Qy 147 SerAspGluGlyAsnIleThr-MetThrSerPhe 157
Db 679 AATGCACGCGGAGTTATTCAGGATGACCTCATTC 712

```

RESULT 21

```

US-09-815-242-6804
; Sequence 6804, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: fastSeq for Windows Version 4.0
; SEQ ID NO 6804
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2157)
US-09-815-242-6804

```

Alignment Scores:

```

Pred. No.: 1-08 Length: 2157
Score: 79.50 Matches: 45
Percent Similarity: 35.84% Conservative: 36
Best Local Similarity: 19.91% Mismatches: 98
Query Match: 7.44% Indels: 47

```

```

; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1945
; LENGTH: 2355
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1945

Alignment Scores:
Pred. No.: 0.577 Length: 2355
Score: 82.00 Matches: 63
Percent Similarity: 37.83% Conservative: 38
Best Local Similarity: 23.60% Mismatches: 80
Query Match: 7.68% Indels: 86
Gaps: 17

US-10-024-955-7 (1-213) x US-09-738-626-1945 (1-2355)
Oy 7 lIeAlaAlaValAlaPheValAlaValSerAlaAspProIle-----HisTyrAspLys 24
Db 172 ATTGCTGTGCTGCTTTTACTATGCTGCCCGCATACATACAGCAAAATGACAG 231
Oy 25 lIeThGlu---GluIleasnLysAlaIlleAspAlaIlleAlaIlleGluGlnSer 43
Db 232 GTGACCTCATGACCATCTCCAAAGATGGTTTCATGAGATTTCACCA----- 276
Oy 44 GluThlIleAspProMetLysValProAspHis---AlaAspLysPheGluArgHisVal 62
Db 277 -----CACAAATTTCCCTGATCATGAGATCTCCACCATCGACCATTTATC 321
Oy 63 GlyIleValAspPheLys---GlyGluLeuAlaMetArgAsn-----IleGluAla 78
Db 322 AATACCGCTGATGATGAAATGCGGATGCAAAATGGTGACCTCCGATTTACATGATGACGCTC 381
Oy 79 ArgGlyLeu-----LysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlu 96
Db 382 CGTGACCTGCTCATAGGTATGATGACGACGAGGTCGACACCGCAACCTGACAGCGTTGAGC 441
Oy 97 GlyIleValLysAlaHisLysLeu-----LeuIleGlyValHisAspIle-----Val 112
Db 442 ATCTTATGATGAATCCATATCGAGCGTTTCATCTCGGTGGATGATGATCAAGACGACT 501
Oy 113 SerMetGlu-----TyrAspLeuAlaTyrLysLeuGlyAspLeuHis 126
Db 502 TCTCTGAGCTGGAATCATCATCTGCTACCTGCTCTGTGATTAAGTATGATGAGCGCTCAT 561
Oy 127 -----ProThrThrHisValIlleSerAspIleGlnAsp-----PheVal 139
Db 562 GCATGCCCGAAGTACCTCATTTATGATGAGGTCCAGACAACTCGGTCTTCGATGCTGTC 621
Oy 140 ValAlaLeu-----SerLeuGluIleSerAspGluGlyAsn 151
Db 622 TTTCGACCTTCGTTTTCGGCGAAAGCACAAATACAGATCTGTACTGTGGTGGTGCCTGCA 681
Oy 152 lIeThMetThrSerPhe----- 157
Db 682 CAGACCTTGATGATGTTCCGCTTCTGTAACCTTAAGGCTCTGAACCTCCCTGGAAGCATC 741
Oy 158 -----GluValArgGln 161
Db 742 GGTGTTTTCGGTACCTACCGTTTTCACCAACCAATTACCGTTCCAAACGAAATTTCTTGAC 801
Oy 162 PheAlaAsnVal-----ValAsnHisIleGlyGlyLeuSerIleLeu 175
Db 802 TTCGCTAATATCCACCTTCGCGATATTGAAGCTTAACCAAGTCCAGAGTATTCAGCTCTTAC 861
Oy 176 AspProIlePheGlyValLeuSerAspValLeuThrAla---lIePheGlnAspThrVal 194
Db 862 GCCAAGCTCTTT-----GATGACACCTACTGCGGACAGCGTTCAAGAAAGAGTT 909
Oy 195 ArgLysGluMetThrLysVal 201
:::|

```

```

Db 910 GAACGTGACATGCACCATGTG 930
RESULT 19
US-09-815-242-7863/C
; Sequence 7863: Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7863
; LENGTH: 2853
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; NAME/KEY: CDS
; LOCATION: (1)...(2853)
US-09-815-242-7863

Alignment Scores:
Pred. No.: 1.2 Length: 2853
Score: 80.50 Matches: 45
Percent Similarity: 39.38% Conservative: 31
Best Local Similarity: 23.32% Mismatches: 72
Query Match: 7.54% Indels: 45
Gaps: 9

US-10-024-955-7 (1-213) x US-09-815-242-7863 (1-2853)
Oy 5 LeuLeuIlleAlaValAlaPheValAlaValSerAlaAspProIleHisTyrAspLys 24
Db 1653 CTTCTGGCCCTGCGCGCGGACGACCAAGCGCGGTGACATGACGCTGTGAACGGAT--- 1597
Oy 25 lIeThGluGluIleasnLysAlaIlleAspAlaIlleAlaIlle----- 40
Db 1596 -----CTGCCGGGTGCGCTCGGGGTTCTTTCACCAAGTGAT 1561
Oy 41 ---GluGlnSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGlu 59
Db 1560 GGTGACGATGATGATCCGCCCGGACCAAGAA---GATATGTGCGAAGCCGGTGAC 1507
Oy 60 ArgHisValGlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArg 79
Db 1506 CACTAGCTGCGTGGGGTGAA-----GGTCTTGAGGAATC 1471
Oy 80 GlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGlu----- 96
|

```

```
^ US-10-024-955-7 (1-213) x US-09-874-923-109 (1-1908)
QY 12 phevalalavalseralasprrllohlshstyrasplysllethrglulgleasnllys 31
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 27 TACTTTATTTAGTACGCCA-----ATCGAGTTTCAACAGCTAACAAATG-GTGAATCTTT 79
QY 32 Alalaleasppalalealalealalegluglinsersgluthrlleaspprometlysval 51
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 ACCGTGCATCCAGTCCGCGAGGTGATGACATACCGACCAAGATCCGGAACATGTCGGT 139
QY 52 Proasphialasplysrshegluatghlsvalglyllevaalspheylyglulgleu 71
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 140 ATTGCTCAGCTGCAC-----CACGGCAAGTCCGACGCTG 172
QY 72 Alamelarganlleglualargglyleuylsglmetylsarglnglinspalaasn 91
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 173 TCCGAGCTCTGCTGTGTGCTGCGGCATCATCAAGATGAGAGAGCGCTGGCATTAAGCGG 232
QY 92 Val-----Lysgllyglugllyllevaalsalansleuileugllyvalhls 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 ATCATGATATACACGCGGATGAG---ATCGCGCGTGT-----ATCAGATCAAG 280
QY 109 Aspspillevalsermetglutyraspleualatyllys---leugllyaspleuhlspro 127
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 281 TCCACCGCCATCTCCATGCACACTACCGCCGACGAGATGATCCGCGATCTGGAT--- 337
QY 128 ThrThrlHisValIleaserasprleaglinsprrhevalValaleuaserleugllyleser 147
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 -----GACCAACAAGCGGACTTCTGATCAACCTG-----ATCGAC 373
QY 148 Aspglugllyasnlllethrmethrserprrhegluvalarglinsprrhealaaasnlvalasn 167
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 374 TCCCGCGGACAGCTCACCTTCACGCTCC---GAGGTGACGCGCGCTTCTGCTGACGAGAC 430
QY 168 Hsllegllyleuaserlleuaspprrllohlshstyrasplysllethrglulgleu 187
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 431 -----GGCGCGCTGCTGTGTGAGCTGCTGAGGCGGTGTGCTGACGAGAGAGACG 484
QY 188 Alalaleprrheglinsprrhevalargylsglmetylsvalleualapro 204
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 485 GTGCTG-----CGCCAGCGCTGACGAGCGCATCCGCTT 520

RESULT 17
US-09-841-132-379
; Sequence 379, Application US/09841332
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 379
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
US-09-841-132-379

Alignment Scores:
Pred. No.: 0.268 Length: 1386
Score: 82.00 Matches: 46
Percent Similarity: 40.00% Conservative: 44
Best Local Similarity: 20.44% Mismatches: 69
Query Match: 7.68% Indels: 66
DB: 10 Gaps: 11

US-10-024-955-7 (1-213) x US-09-841-132-379 (1-1386)
QY 23 Asplysillethrglulgu-----lleasnllysalaleasppalaleala 38
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 AATTAATTTTACGAACAAGGAATTCGAATTTCTTACCAAGCC-----TGCATCTCT 705
QY 39 Alalegluglinsersgluthrlleaspprometlysvalproasphialsasplys 58
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 706 GCAATGCAAGAAATCCCAAAACCAAGTTCCGATTACTGTGACAGCATCAAGTGAAGAGTTT 765
QY 59 Glu-----Argnlhs 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 766 GATTATGCTTGTGCTGCTATTTGCTCCCAATTTAATACACCAAGTATAGGCTAGATAT 825
QY 62 Valglylleval---Asprrheylslyglullevaalemararganlleglualarggly 80
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 826 GCTGAGATGATCCGGAGACGATGCTGGCGGTGATTCGTTGACCAAAACCATGCCCATAT 885
QY 81 Leuylsglmetylsarglngllyasr-----Alaasnvallys 93
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 886 GTTCCAAATATCTATGCGATTGGAGACATCACTGAAAGTGGCTACTTGTCTATGTGCT 945
QY 94 GlyglugllyllevallysAlansleuileglyvalhlsaspspillevalser 113
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 946 TCCGACCAAGCGGTATTTGCCGCAAAATATTTGCGGACATACGAA-----GTT 996
QY 114 MetglutyraspleualatyllysleugllyaspleuhlsproThrlHisValIleaser 133
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 997 ATGATATTATCTGCCATACCTTCTGTGATCTTACCACCA----- 1038
QY 134 AsplleglinsprrhevalValaleuaserleugllyleseraspugllyasnle--- 152
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1039 -----GAAATTCCTATGAGCTCTATCTCTCAAGAAAGACAGAACAAATCTTCC 1092
QY 153 ---Thrmethrserprrhegluvalargln----- 161
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1093 GCAAACTCCCAAAATTCCTTTAAAGCATTGAAAGCTGTCTTGGAGACATCT 1152
QY 162 -----PhealasnvalValanshl-----llegllyleuaserlle 174
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1153 GATGCTTTGCTCTATTTGTGAGTCAATGAATTAACCAAGAAATTAACGAGTATATGTC 1212
QY 175 LeuaspprrllohlshstyrasplyslleuaserasppvalleuThrlalaleprrheglinsprrheval 194
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1213 ATAGACCTTCACGCTCATCATCAATTAATTTGAGAGATGACCTTA-----GCGATC 1260
QY 195 Argylsglmethr 199
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1261 CGCAATGAGCTGAC 1275

RESULT 18
US-09-738-626-1945
; Sequence 1945, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
```



```

US-10-024-955-7 (1-213) x US-09-738-626-1 (1-3309400)
QY 54 HisAlaAspIleuValArgHisValGlyIleValAlaSerPheLeuGlyGluLeuAlaMet 73
Db 22563 TACGGCGAAGCTTTTGA-----ATCCAGCCTTTTCGCGAATGATCTATT 22610
QY 74 ArgAsnIleGluValArgGlyLeuGlyGluMetLysArgGlnGlnLysAspAlaAsnValLys 93
Db 22611 GTACCCCTCAAGGATCAAGCTTCMAAGATGCCCTTGAGAG-----CAG 22655
QY 94 GlyGluGluGlyIleValLysAlaHisLeuLeuIleGlyValHisAspIleValSer 113
Db 22656 TGGGAGAAGAGTTCTGCACACAGTCGACACTTGCCTATCCGACAC---GTTTCC 22712
QY 114 MetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIleSer 133
Db 22713 TACACCTACCAATCAACCTCCATCGGTGAC---CGCTCACTTCCGACCATTTGAT 22769
QY 134 AspIle-----GlnAspPheValAlaLeuSerLeuGluIleSerAsp 148
Db 22770 GATACCCCACTTGATCCGAGACGACTGTTGTCACACTTCCCTGACTC---CAG 22826
QY 149 GluGlyAsnIleThrMetThrSer 156
Db 22827 TCCGCAACGAGATGATGACCGCA 22850

RESULT 15
US-09-991-496-109
; Sequence 109, Application US/09991496
; Patent No. US20020169285A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Melo, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Colier, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C9
; CURRENT APPLICATION NUMBER: US/09/991,496
; CURRENT FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 109
; LENGTH: 1908
; TYPE: DNA
; ORGANISM: Leishmania major
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1738, 1855, 1864, 1879, 1882, 1903
; OTHER INFORMATION: n = A,T,C or G
US-09-991-496-109

Alignment Scores:
Pred. No.: 0.314 Length: 1908
Score: 83.00 Matches: 47
Percent Similarity: 43.65% Conservative: 39
Best Local Similarity: 23.86% Mismatches: 75
Query Match: 7.77% Indels: 37
Db: 9 Gaps: 11

US-10-024-955-7 (1-213) x US-09-991-496-109 (1-1908)
QY 12 PheValAlaValSerAlaProIleHisTyrAspLysIleThrGluGlnIleAsnLys 31
Db 27 TACTTTATTTCTAAGCGCA-----ATCGACTTTCACACAGCTAACAATAAGT-GTGAACTTT 79
QY 32 AlaIleAspAspAlaIleAlaAlaIleGluGlnSerGluThrIleAspPrometLysVal 51

```

```

Db 80 ACCGTGATCAGGTCCGCGAGGTAGTGAATCCGACAGATCCGGAACATGTCCTG 139
QY 52 ProAspHisAlaAspLysPheGluValArgHisValGlyIleValAspPheLysGlyLeu 71
Db 140 ATTGTCTTCTGCTG-----CAGCGCAAGTCGACGCTG 172
QY 72 AlaMetArgAsnIleGluValArgGlyLeuGlyGluMetLysArgGlnGlnLysAspAlaAsn 91
Db 173 TCCGACTCTCTGCTGTGCTGCTGCCGCATCATCAATGATGAGAGGCTGGCATTAAGCGG 232
QY 92 Val-----LysGlyGluGluGlyIleValLysAlaHisLeuLeuIleGlyValHis 108
Db 233 ATCATGATATACAGCCGCGATGAG---ATCGCGCTGGT-----ATCAGATCAAG 280
QY 109 AspAspIleValSerMetGluTyrAspLeuAlaTyrLys-----LeuLysAspLeuHisPro 127
Db 281 TCCACCGCCCATCTCCATGACACTACACAGTCGCGAAGAAATGATGCGCATGTGAT--- 337
QY 128 ThrThrHisValIleSerAspIleGlnAspPheValAlaLeuSerLeuGluIleSer 147
Db 338 -----GACGACAAAGCCGACTTCTGATCAACTG-----ATCGAC 373
QY 148 AspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValAlaAsn 167
Db 374 TCCCGCGACAGCTGACACTTCACCTCC---GAGTGATCTCCGCTCTGCTGAGACGAC 430
QY 168 HisIleGlyLysLeuSerIleLeuAspProIlePheGlyValLeuSerAspValIleThr 187
Db 431 -----GGCGCGCTGTGCTGTGAGCTGCTGAGAGGCGCTGTGCTGACAGCGAGACG 484
QY 188 AlaIlePheGlnAspThrValArgLysGluMetThrLysValLeuAlaPro 204
Db 485 GTGCTG-----CGCCAGCGCGCTACCGAGCGCATCGCGCT 520

RESULT 16
US-09-874-923-109
; Sequence 109, Application US/09874923
; Patent No. US20020081320A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Melo, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Colier, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 109
; LENGTH: 1908
; TYPE: DNA
; ORGANISM: Leishmania major
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)---(1908)
; OTHER INFORMATION: n = A,T,C or G
US-09-874-923-109

Alignment Scores:
Pred. No.: 0.314 Length: 1908
Score: 83.00 Matches: 47
Percent Similarity: 43.65% Conservative: 39
Best Local Similarity: 23.86% Mismatches: 75
Query Match: 7.77% Indels: 37
Db: 10 Gaps: 11

```

```
Alignment Scores:
Pred. No.: 0.0286 Length: 406
Score: 83.50 Matches: 30
Percent Similarity: 43.44% Conservative: 23
Best Local Similarity: 24.59% Mismatches: 52
Query Match: 7.82% Indels: 17
DB: 10 Gaps: 3

US-10-024-955-7 (1-213) x US-09-960-352-7528 (1-406)
QY 18 ASPPROILEHISTYRASPRLYSILETHRGUGLILAEALNLYSALAILEASPSPALAILIE 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17 GACAGSTTTCCTTCGACCTCGCGAAGAGGAGGATCATCAGGCTTGATATTCCTGTA 76
QY 38 AIAAIAIEGLUGLINSERLURHILE----- 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77 GCGACCATGAGGTGGGTGAAGTGTCCATATCAGCTGCAGACCCAGATACGCTATGCG 136
QY 47 -----ASPPrometLysValProAspHisAlaAsp---LysPheGLUArgHisVal 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 137 TTAGCGGAGGAGCCCCCGAAGATCCCTCCCAAGCCACGCTTGTTGAG-----GTG 190
QY 63 GLYLIEVALASPPhelyslygluleuAlaMetArgAsnILEGLUALArgglyLeuLys 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 191 GAGCTGTTCAGTTCAGGAGGTGAGACTGACAGAGAGGAGGAGCGCGAATCATCAGG 250
QY 83 GlnMetLysArgGLUGLysAlaAsnValLysGLUGLUGLysLysValLysAlaHis 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 251 AGAATACGAGCGCGGGGAGGCTATGCCAAGCCCAACAGAGTCTCTCGTGAGGTT 310
QY 103 LeuLeuILEGLYValHisAspRIleValIserMetGLUTyrAspLeuAlaYrLysLeu 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 GCACGTGAGAGATATTCTCAAGACCAAGGTGTTTGACCGGCGGAGCTCCGCTTGAGTTC 370
QY 123 GLYASP 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 371 GCGGAA 376

RESULT 13
US-09-738-626-26
; Sequence 26, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 26
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-26

Alignment Scores:
Pred. No.: 1.34e+04 Length: 3309400
Score: 83.50 Matches: 34
Percent Similarity: 46.30% Conservative: 16
Best Local Similarity: 31.48% Mismatches: 41
Query Match: 7.82% Indels: 17
DB: 9 Gaps: 6
```

```
Pred. No.: 0.144 Length: 1236
Score: 83.50 Matches: 34
Percent Similarity: 46.30% Conservative: 16
Best Local Similarity: 31.48% Mismatches: 41
Query Match: 7.82% Indels: 17
DB: 9 Gaps: 6

US-10-024-955-7 (1-213) x US-09-738-626-26 (1-1236)
QY 54 HISAlaAspLysPheGLUArgHisValIGLYILEVALASPPhelyslygluleuAlaMet 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 400 TAGCGCGAGGCTTTGAA-----ATCCAGCCTTCTCCGGTGAATTCATTG 447
QY 74 ArgAsnILEGLUALArgglyLeuLysGlnMetLysArgGLUGLYAspAlaAsnValLys 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 448 GTCACCCCTCAAGGATCAACACTTCACAAAGATGCCCTTGACAG-----CAG 492
QY 94 GLUGLUGLUGLYILEVALysAlaHisLeuLeuILEGLYValHisAspRIleValLys 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 493 TGGGAAGAGGTTCTGCACGACAGTGCAGCACTTGCGGTATCCGACAAAC---GTTTCC 549
QY 114 MetGLUTyrAspLeuAlaYrLysLeuGLYAspRIleHisProThrHisValLysSer 133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 550 TACACCTACAGATCAACCGTCCAAATCGGTGAC---CGCGTCACTTCGTCACCATTGAT 606
QY 134 AsPIle-----GlnAspPheValValAlaLeuSerLeuGlnLysSerAsp 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 GATACCCCACTGATCCGGAACGACGACATACGTTTGCGAGCTTCCTGTACCTC---CAG 663
QY 149 GLUGLYAsnILEThrMetThrSer 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 664 TCCGCAACGAAGTATGACCGCA 687

RESULT 14
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.: 1.34e+04 Length: 3309400
Score: 83.50 Matches: 34
Percent Similarity: 46.30% Conservative: 16
Best Local Similarity: 31.48% Mismatches: 41
Query Match: 7.82% Indels: 17
DB: 9 Gaps: 6
```

QY 41 GIUGInSerGIuThrIleAspPronMetLysVal----- 51
|||
Db 1171 GAGCAAAATATTCATATAAAACAATCGAATAATTCCTAGAGATTATAGTAT 1230
QY 52 -----ProAspHisAlaSprLysPheGluArg 60
Db 1231 AATATCTCGAAAGAAATTTTAAGACAACTCTTAATACAGGTGAACGTGTGAACGT 1290
QY 61 -----HisValGIyLLeValAspPheLysGIy----- 69
|||
Db 1291 GGTGACAGCTTGATGTCTGTATATCAAAAGCCCTGAAAGGTTAAATGCAATATGC 1350
QY 70 -----GLeuLeuAlaMetArgAsnIleGluAlaArgGIyLeuLysGIuMet 84
|||
Db 1351 ATTTGGTTTACTTAAGAGCAACCTTGACAAATTTAAATCGTTAGCTTTAAAGATGTT 1410
QY 85 LysArgGIuGIyAspAlaAsnValLysGIyGIuGIuGIyIle-----Val 99
|||
Db 1411 AAGATTGAAAAGTATATATATATCAAGCCCAAGGATACATTGCAAACTGA 1470
QY 100 LysAlaHisLeuLeuIleGIyValHisAspAspIleValSerMetGIuThyAspLeuAla 119
|||
Db 1471 ACCGCAATACGCAATCCCTATCTCATGTCTTAATATTAACATATGAATCTTTAGGC 1530
QY 120 TyrLys 121
|||
Db 1531 ATTAAG 1536

RESULT 11
US-09-815-242-7737/c
; Sequence 7737, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7737
; LENGTH: 2832
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2832)
US-09-815-242-7737

Alignment Scores: 0.262 Length: 2832
Pred. No.:

Score: 85.50 Matches: 55
Percent Similarity: 37.96% Conservative: 27
Best Local Similarity: 25.46% Mismatches: 106
Query Match: 8.01% Indels: 28
Gaps: 8
DB: 10

US-10-024-955-7 (1-213) x US-09-815-242-7737 (1-2832)

QY 6 LeuIleAlaAlaValAlaPheValAlaValSerAlaAspProIleHisTyrAspLysIle 25
|||
Db 1404 CTCGTTGTGGCGCGCGGACGAGTACAGACAGTCGAT-----GACCACGTC 1357
QY 26 ThrGIuGIuIleAsnLysAlaIleAspAlaIleAlaAlaIleGIuInSerGIuThr 45
|||
Db 1356 ACCTTGAACCTCATCCGCTAGTACGACGCCATCTGGTGAGCAACACAGACGCTTCGGC 1297
QY 46 IleAspPronMetLysValProAspHisAlaSprLysPheGluArgHisValGIyIleVal 65
|||
Db 1296 ATCTGTCGCGGTGAGCTGGAAGATGCGCGCTGAATCATCTTCCGACGTCAGT----- 1243
QY 66 AspPheLysGIyLLeuLeuAlaMetArgAsnIleGluAlaArgGIyLeuLysGIuMetLys 85
|||
Db 1242 -----GCGTACTCGGTGAGACGCCATCTGCTGCGCGGTGTAACCGAC 1192
QY 86 ArgGIuGIyAspAlaAsnValLysGIyGIuGIuGIyIleValLysAlaHisLeu----- 103
|||
Db 1191 CTGGTTGTTGATCACGAGTGTGATGTCACCGGCTTGTAGCGCGGTCTCGACAT 1132
QY 104 LeuIleGIyValHisAspAspIleValSerMetGIuThyAspLeuAlaTyrLysLeuGIy 123
|||
Db 1131 CTGGAAGTTTCCATGACACCCGCTGCGCGCAAGGCGGATGCGCTGGATGGAAT 1072
QY 124 AspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValAlaLeuSer 143
|||
Db 1071 CGGTACCTCTGTGCGCGGACCTGTC-----CTTGGCGGCTCTGCGGACGACG 1018
QY 144 LeuGIuIleSerAspGIuGIyAsnIleThrMetThrSerPheGluVal----- 159
|||
Db 1017 GGAACCTCGACACCGCGGAAC-----GATTTCCAGGTGGAGCGGTGGA 970
QY 160 -----ArgGIu-----PheAlaAsnValAlaHisIleGIyIle-----LeuSer 173
|||
Db 969 CGCCAGCGCGGAGTGTGACCTCCACCGCTGTCATGACGTTGAGGACGACCCCTGTG 910
QY 174 IleLeuAspProIlePheValIleuSerAspValIleThrAlaIlePheGlnAspThr 193
|||
Db 909 GTACTTCACGTACCGGACCCGACGCTGACGATCTTCTTCCCTTGACCTGCGACAG 850
QY 194 Val-ArgLysGIuMetThrLysValLeuAlaProAlaPheLysArg 208
|||
Db 849 GTCCGGCGGCTTCTTGCACGAGGTGTGACGAGGCTTCAGCGCA 804

RESULT 12
US-09-960-352-7528
; Sequence 7528, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIOR FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7528
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 32-LIB34-049-Q1-E1-H7
US-09-960-352-7528

```

US-09-815-242-4415
; Sequence 4415, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4415
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4415

Alignment Scores:
Pred. No.: 0 115 Length: 1980
Score: 86.50 Matches: 36
Percent Similarity: 41.36% Conservative: 31
Best Local Similarity: 22.22% Mismatches: 50
Query Match: 8.10% Indels: 45
DB: 10 Gaps: 5

US-10-024-955-7 (1-213) x US-09-815-242-4415 (1-1980)
QY 1 MetMetLysPheLeuLeuIleAlaIaValAlaPheValAlaValSerAlaAspProIle 20
DB 1060 ATCTTTGCTGTATATGATTCACACTGTCTTTGTCGCAATGCAATGTTGGTAAT 1119
QY 21 HisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAlaIleAlaIle 40
DB 1120 AAATACGAAGAAGACACGATGTAATCGGAATCTGTAAAGAAGCA----- 1167
QY 41 GluGlnSerGluThrIleAspPrometLysVal----- 51
DB 1168 GAGCAAAATATTCATAAACAACATCTGAAATTAAGGTAAATTCTAGAGTTATAGTGAT 1227
QY 52 -----ProAspHisAlaAspLysPheGluArg 60
DB 1228 AAATATCTCGAAATATTTTAAGACAACTCCTAATACAGGTGAACGTGTGAACGT 1287
QY 61 -----HisValGlyIleValAspPheLysGly----- 69
DB 1288 GGTGACAGAGTGTGATGTATATCAAAAGCCCTGAAAGAGTTAAATGCCAATGTC 1347
QY 70 -----GluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMet 84
DB 1348 ATTGGTTTAACTTAAGAGCAACGCTTGACGAATTAATTAATCGTTAGTCTTAAAGATGTT 1407

```

```

QY 85 LysArgGlnGlyAspAlaAsnValLysGlyGlnGlyIle-----Val 99
DB 1408 AAGATTGAAAGATATATATATATCAAGGCCAAGATACATTGCAATGCAAGTGA 1467
QY 100 LysAlaHisLeuLeuIleGlyValHisAspAlleValSerMetGluTyrAspLeuAla 119
DB 1468 ACCGCAAAATCTGAAATCGCATATCATCTTAATTAATTAATGAATCTTAAGGC 1527
QY 120 TyrLys 121
DB 1528 ATTAAG 1533

RESULT 10
US-09-815-242-8081
; Sequence 8081, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8081
; LENGTH: 1995
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1995)
US-09-815-242-8081

Alignment Scores:
Pred. No.: 0 117 Length: 1995
Score: 86.50 Matches: 36
Percent Similarity: 41.36% Conservative: 31
Best Local Similarity: 22.22% Mismatches: 50
Query Match: 8.10% Indels: 45
DB: 10 Gaps: 5

US-10-024-955-7 (1-213) x US-09-815-242-8081 (1-1995)
QY 1 MetMetLysPheLeuLeuIleAlaIaValAlaPheValAlaValSerAlaAspProIle 20
DB 1063 ATCTTTGCTGTATATGATTCACACTGTCTTTGTCGCAATGCAATGTTGGTAAT 1122
QY 21 HisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAlaIleAlaIle 40
DB 1123 AAATACGAAGAAGACACGATGTAATCGGAATCTGTAAAGAAGCA----- 1170

```

```

; TYPE: DNA
; ORGANISM: P. falciparum
; US-09-742-096-1

Alignment Scores:
Pred. No.: 0.207 Length: 6152
Score: 90.00 Matches: 47
Percent Similarity: 44.55% Conservative: 43
Best Local Similarity: 23.27% Mismatches: 78
Query Match: 8.43% Indels: 34
DB: 9 Gaps: 9

US-10-024-955-7 (1-213) x US-09-742-096-1 (1-6152)
OY 22 TyrAspLysIleHrGluGluIleAsnLysAlaIleAspAlaIleAlaIleGlu 41
DB 3179 TTTAAATCTGTATTAGTAAAGTAAGCAAGTACAGAAATTTACCGAAGATTGAA 3238
OY 42 GlnSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGluArgHis 61
DB 3239 AACAAATGAA-----ATCGATAAAGCATTTTGTGTAATATTTGATTAATGTA 3286
OY 62 ValGlyIleValAspPheLysGlyIleValAlaMetArgAsnIleGluAlaArgGlyLeu 81
DB 3287 AAGGAATACAGAAATTTATTAAACAGATGTTTCAAGATATAGAACCAAGATATGTA 3346
OY 82 LysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyIleValLysAla 101
DB 3347 ATCCAAATCAGAAAGAAAGGTGATTTGAAT-----GAAATGTGGTTAGTTCG 3394
OY 102 HisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLys 121
DB 3395 ATTTTA-----GATAATATAGAAATATGAAAGAAAGCTTTATTAAATGAA 3439
OY 122 LeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValAla 141
DB 3440 TTGAAATATTTTCAAGTCAAGGTTCAGAAACCTGTAACCTGAACATGTA----- 3493
OY 142 LeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluVal----- 159
DB 3494 -----GAAACAAATGTA---TATGTGATGTTGATGTTCCCTGCT 3529
OY 160 -----ArgGlnPheAlaAsnValValAsnHisIleGlyLysLeuSer----- 173
DB 3530 ATGAAGATCATATTTTAAAGATATTAATGAGCAGAGGTTGAAGAAATGTTTTT 3589
OY 174 IleLeuAspProIlePheGlyValLeuSerAspValLeuThrAla-----IlePheGln 191
DB 3590 AATTGGACAGATGATTTTAAAGTGAAGATGATGATGATGATGATGATGATGATGAT 3649
OY 192 AspThrValArgLysGlnMetThrLysValLeuAlaProAlaPheLysArgGluLeuGlu 211
DB 3650 GAACCGGTTAAAGAGGTGAGAAAGAAAGAACTGTTAGTATTGTA---GAAATGGA 3706
OY 212 LysAsn 213
DB 3707 GAAAT 3712

RESULT 8
US-09-796-692-8801
; Sequence 8801, Application us/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01

```

```

; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 8801
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (524)
; OTHER INFORMATION: n=A,T,C or G

US-09-796-692-8801

Alignment Scores:
Pred. No.: 0.0113 Length: 607
Score: 88.50 Matches: 30
Percent Similarity: 45.08% Conservative: 25
Best Local Similarity: 24.59% Mismatches: 50
Query Match: 8.29% Indels: 17
DB: 9 Gaps: 3

US-10-024-955-7 (1-213) x US-09-796-692-8801 (1-607)
OY 18 AspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAlaIle 37
DB 76 GACAAATTTCTCTTGAACCTGGGAAAGGGAGGTGATCAAGCTTGGACATGGCATA 135
OY 38 AlaAlaIleGluGlnSerGluThrIle----- 46
DB 136 GCCACCATGAAGGTGGGGAGGTGTGCCACATGCACCTGCAACGACGATATGCGTACG 195
OY 47 -----AspProMetLysValProAspHisAlaAsp---LysPheGluArgHisVal 62
DB 196 TCAGCAGCAGCAGCTCCCAAGATTTCCCAATGCCACGCTGATTTGAG-----GTG 249
OY 63 GlyIleValAspPheLysGlyIleValAlaMetArgAsnIleGluAlaArgGlyLeuLys 82
DB 250 GAGCTGTTTGAAGTTTAAAGGAGATGTGACGGAAGAGAGAGAGAGAGAGAGAGAGAG 309
OY 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLysAlaHis 102
DB 310 AGAATACAGACTCGCGGTGAAGGCTATGCTACACCCCAATGAGGTTGATGTCGAGAGGT 369
OY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
DB 370 GCACCTGGAGGATACAGAGCAAGAGCTCTTGACACGCGGAGGCTCCCTTTGAGATT 429
OY 123 GlyAsp 124
DB 430 GCGGAG 435

RESULT 9

```



```

; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: 15966-776CIP
; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/215,855
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/839,446
; PRIOR FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: POLYX
; OTHER INFORMATION: CG55655_02
US-09-898-570-31

Alignment Scores:
Pred. No.: 0.00205 Length: 892
Score: 96.00 Matches: 44
Percent Similarity: 40.54% Conservative: 31
Best Local Similarity: 23.78% Mismatches: 58
Query Match: 8.99% Indels: 52
DB: Gaps: 10

US-10-024-955-7 (1-213) x US-09-898-570-31 (1-892)
Oy 27 GUGLUtLeAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIle 46
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 310 GAAATACATCAAGAGAGATTGATGATTTAGTTAAAGAGTTAAAGTCAGAGTTGAA 369
Oy 47 Asp---PrometLysVal-----ProAspHisAlaAspLysPheGlu 59
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 AATGTCATCTTCAGTCAAGAGATCTTAATCTCAGATGCTGCAATGTTTC--- 426
Oy 60 ArgHisValGlyIleValAspPhe-----LysGlyGlu--- 70
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 CGCCATTTTCAGCAAAATCATGTTTATATACATGACACAAATAGCACAAGAGAA 486
Oy 71 -----LeuAlaMetLArgAsnIleGluAlaArgGly----- 80
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 TGCAGACATTTATTTTACGTCAAGTTGAAAGTTGCGGAAGAGATGCTGAAGAAT 546
Oy 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGlu----- 96
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 GTAATATGATATGCTTCATCAAGAAATGGAAGTTTAAATGAAGCTTACTTACAGAA 606
Oy 97 ---GlyIleValLysAlaHisIleu-----LeuIleGlyValHisAspIleValSer 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 ATCAATATCATTAAGCAACACTTTCAGAGATTGAACAGACAGACAAAGAACTTGTAAAT 666
```

```

Oy 114 MetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrHisValIleSer 133
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 TTGGAG-----AACCAATTAAG 684
Oy 134 AspIleGlnAspPheValAlaIleuSerLeuGluIleSerAspGluIly----- 150
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 685 GATTTAAGGATCTTTCTTCAGATATCTCTTTAGTAGAGAGAACAGAGAGCATTC 744
Oy 151 ---AsnIleThrMetSerPheGlnValArgGlnPheAlaAsnValAlaHisIle 169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 745 AACCAATTTGAATGACAGATGATACAAAGAGTATGTTTAACATATCTAAAGAAA 804
Oy 170 GlyGlyLeuSerIle 174
    |||||:|||||
Db 805 TTTGACACTAGCTGTA 819

RESULT 5
US-09-898-570-27
; Sequence 27, Application US/09898570
; Patent No. US20020123612A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELBERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; FILE REFERENCE: 15966-776CIP
; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 09/839,446
; PRIOR FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 893
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: POLYX
; OTHER INFORMATION: h_nh0778p17_A
US-09-898-570-27

Alignment Scores:
Pred. No.: 0.00206 Length: 893
Score: 96.00 Matches: 44
Percent Similarity: 40.54% Conservative: 31
Best Local Similarity: 23.78% Mismatches: 58
Query Match: 8.99% Indels: 52
DB: Gaps: 10
```

```

QY      41  GlnuInserGIurhrlleAspProwetLysValProAspHisAlaAspLysPheGluArg 60
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      188  GAAAAATCCGAACATTCGATCCAAATGAAGGTRACCCGATCATTCGATAAATCGAAGCA 247
QY      61  HisValGlylleValAspPheLysGlyLysLeuAlaMetArgAsnIleGluAlaArgGly 80
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      248  CATTTGCGTATCATTCGATTTTAAAGGTGAATTAAGACATCGCAAAACATTCAGTTGAGAGA 307
QY      81  LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys 100
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      308  TTTAAACAAATGAAACGTGTAGCTGATGCTAAATGTGAAAGTGAAGATGTGCTGCAAA 367
QY      101  AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluLysAspLeuAlaLyr 120
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      368  GCTCATTTGTTGGTCGGGTGTTCTATGATGACGTTGTTCAATGGAATATGATTTAGCAATAC 427
QY      121  LysLeuGlyAspLeuHisProThrTrpHisValIleSerAspIleGluAspPheValVal 140
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      428  AAATGGCGTATCTTCATCCAAACACTCATGTCATTTCCGATATTCAGAGATTTGTTGTC 487
QY      141  AlaLeuSerLeuGluLysSerAspGluGlyAsnIlePheMetLysSerPheGluValArg 160
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      488  GAATTCATCGCTCAAGATTCAGCAAGAAGTAATATGACATTCGATCGATCGAAGTACGT 547
QY      161  GlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheLys 180
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      548  CAAATTCGCAATGTGTCAATCATATATGGTGTCTTCAATTTTGGATCCAAATTTGCTCT 607
QY      181  ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetLys 200
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      608  GCTCATTCGATGTTTGGACCGCGTATTTCCAGAGATACCGTACGTGACAGAAATGACCAA 667
QY      201  ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
      |||:::|||||  |||:::|||||:::|||||:::|||||
Db      668  GTATTGCGACGACATTCAAAAAAGAATTGGAACGCAAC 706

RESULT 3
US-09-898-570-29
: Sequence 29, Application US/09898570
: Patent No. US20020123612A1
: GENERAL INFORMATION:
: APPLICANT: GERLACH, VALERIE L.
: APPLICANT: ELLERMAN, KAREN
: APPLICANT: MACDOUGALL, JOHN R.
: APPLICANT: SMITHSON, GLENDA
: TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
: FILE REFERENCE: 15966-776CIP
: CURRENT APPLICATION NUMBER: US/09/898,570
: FILING DATE: 2001-07-03
: PRIOR APPLICATION NUMBER: 60/198,293
: FILING DATE: 2000-04-19
: PRIOR APPLICATION NUMBER: 60/198,645
: FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: 60/210,809
: FILING DATE: 2000-06-09
: PRIOR APPLICATION NUMBER: 60/199,476
: FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/200,025
: FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/224,610
: FILING DATE: 2000-08-11
: PRIOR APPLICATION NUMBER: 60/200,024
: FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/199,880
: FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/218,591
: FILING DATE: 2000-07-17
: PRIOR APPLICATION NUMBER: 60/271,814
: FILING DATE: 2001-02-27
: PRIOR APPLICATION NUMBER: 60/215,855
: FILING DATE: 2000-07-03

```

```

: PRIOR APPLICATION NUMBER: 09/339,446
: PRIOR FILING DATE: 2001-04-19
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 29
: LENGTH: 892
: TYPE: DNA
: ORGANISM: Unknown Organism
: FEATURE:
: OTHER INFORMATION: Description of Unknown Organism: POLYX
: OTHER INFORMATION: hnh078p17_A1
US-09-898-570-29

Alignment Scores:
Pred. No.: 0.00205 Length: 892
Score: 96.00 Matches: 44
Percent Similarity: 40.54% Conservative: 31
Best Local Similarity: 23.78% Mismatches: 58
Query Match: 8.99% Indels: 52
DB: 10 Gaps: 10

US-10-024-955-7 (1-213) x US-09-898-570-29 (1-892)
QY 27 GUGUUIleAsnLYsAlaileAspAlaileAlaileGLUGInserLUthrlle 46
||| |||||:::|||||::: |||
Db 310 GAATATCATCAACAGAAAGTTGATGATTTAGTTAAAGAAGTTAAAAAGTCAGAGCTTCAA 369
QY 47 Asp--PrometLYsVal-----ProAsPHisAlaAspLYsPheGLu 59
::: ||| ||| |||||
Db 370 AATGGTCATCTTCAGGTGCACAGAGATCTTAATATCTCAGCATGTGCATGTTCTC-- 426
QY 60 ARGHsValGLYleValAspHe-----LYsGLYlu-- 70
||| ||| ::| |||
Db 427 CGCCATTTTCAGCAAAATCATGTTTATATACATAGACACATAGCAGCAAAAGCAAGAGAAG 486
||| ||| ::| |||
QY 71 -----LeuAlaMeLArGAsnIleGLuAlaArgLY----- 80
||| ||| ::| |||
Db 487 TGCAGACATTTATTATTACGTCAGCTTGAGATTGCTGGAAAAAGAGATGCTGAGAGAGAT 546
||| ||| ::| |||
QY 81 LeuLYsGLInMeLYsArGInGLYAspAlaAsnValLYsGLYGLu----- 96
||| ||| ||| |||
Db 547 GTAATGATATGCTTCATCAAGGAATAAGTGAAGTTTAAATGAAGAAGCTTACTTACAGAA 606
||| ||| ||| |||
QY 97 ---GLYleValLYsAlaHIsLeu-----LeuIleGLYValHIsAspSPleValSer 113
||| ||| ||| ||| ::| |||
Db 607 ATCAATATCACTAAGCAACAACTTTCAGAGATTGAACAGACAGACAGAAAGAACTGTATAT 666
||| ||| ||| ||| ::| |||
QY 114 MetGLUtyrAspLeuAlaTYrLYsLeuGLYAspLeuHIsProTHrHIsValIleSer 133
::| ||| ::| |||
Db 667 TTGGAG-----AACCAATTAAG 684
||| |||
QY 134 AsPIleGLInAspPheValAlaLeuSerLeuGLYIleSerAspGLYlu----- 150
||| ||| ::| |||
Db 665 GATTTAAAGGAGTCTTTTCATTCAGATATCTCTTTTGTAGAGAAACAAAGAGAGAGCATC 744
||| ||| ||| |||
QY 151 --AsnIleThrMetHrSerPheGLUValArgInPheAlaAsnValValaSnHIsIle 169
||| ||| ||| ||| ::| |||
Db 745 AACCATATATGAATAGACAGTCAATAGTACAAAGAAAGAACTATTGTTAAACATACTAAAGAAA 804
||| ||| ||| |||
QY 170 GLYleuSerIle 174
||| ||| ||| |||
Db 805 TTTGGACTAGCTGTA 819
||| ||| ||| |||

RESULT 4
US-09-898-570-31
: Sequence 31, Application US/09898570
: Patent No. US20020123612A1
: GENERAL INFORMATION:
: APPLICANT: GERLACH, VALERIE L.
: APPLICANT: ELLERMAN, KAREN
: APPLICANT: MACDOUGALL, JOHN R.
: APPLICANT: SMITHSON, GLENDNA
: TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 17:39:19 ; Search time 118 Seconds
(without alignments)
919.372 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 1068
Sequence: 1 MMKFLIAAFAVAVSADPI.....VRKEMTKVLAPFRLEKN 213

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 424239 seqs, 254661826 residues
Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPO.spool/US10024955/runat_14022003.130700.23811/app.query.fasta_1.391
-DB=Published_Applications_NA -OFM=fastlap -SUFFIX=trpb -MINMATCH=0.1
-DOOPCL=0 -DOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=Dlosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcpt -THR_MAX=100
-THR_MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10024955.qcgn1.1.80 @runat.14022003.130700.23811
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_NMAP -LARGEDUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCCT_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1068	100.0	761	9	US-10-024-955-6	Sequence 6, Appli
2	962	90.1	812	9	US-10-024-955-1	Sequence 1, Appli
3	96	9.0	892	10	US-09-898-570-29	Sequence 29, Appli
4	96	9.0	892	10	US-09-898-570-31	Sequence 31, Appli

5	96	9.0	893	10	US-09-898-570-27	Sequence 27, Appli
6	90	8.4	5361	9	US-09-742-096-2	Sequence 2, Appli
7	90	8.4	6152	9	US-09-742-096-1	Sequence 1, Appli
8	88.5	8.3	607	9	US-09-766-692-8801	Sequence 8801, Ap
9	86.5	8.1	1980	10	US-09-815-242-4415	Sequence 4415, Ap
10	86.5	8.1	1995	10	US-09-815-242-8081	Sequence 8081, Ap
11	85.5	8.0	2832	10	US-09-815-242-7737	Sequence 7737, Ap
12	83.5	7.8	406	10	US-09-960-352-7528	Sequence 7528, Ap
13	83.5	7.8	1236	9	US-09-738-626-26	Sequence 26, Appli
14	83.5	7.8	309400	9	US-09-738-626-1	Sequence 1, Appli
15	83	7.8	1908	9	US-09-991-456-109	Sequence 109, App
16	83	7.8	1908	10	US-09-874-923-109	Sequence 109, App
17	82	7.7	1386	10	US-09-841-132-379	Sequence 379, App
18	82	7.7	2355	9	US-09-738-626-1945	Sequence 1945, Ap
19	80.5	7.5	2853	10	US-09-815-242-7863	Sequence 7863, Ap
20	80	7.5	2493	9	US-09-738-626-1968	Sequence 1968, Ap
21	79.5	7.4	2157	10	US-09-815-242-6804	Sequence 6804, Ap
22	79.5	7.4	18946	9	US-10-147-026-15	Sequence 15, Appli
23	79	7.4	1650	9	US-09-738-626-2243	Sequence 2343, Ap
24	79	7.4	20072	10	US-09-070-927A-89	Sequence 89, Appli
25	78.5	7.4	1680	10	US-09-815-242-7806	Sequence 7806, Ap
26	78.5	7.4	2430	10	US-09-815-242-4531	Sequence 4531, Ap
27	78.5	7.4	2457	10	US-09-815-242-8326	Sequence 8326, Ap
28	78.5	7.4	2457	10	US-09-815-242-8691	Sequence 8691, Ap
29	78.5	7.4	3627	9	US-09-738-626-1074	Sequence 1074, Ap
30	78.5	7.4	32768	10	US-09-070-927A-17	Sequence 17, Appli
31	78	7.3	648	10	US-09-815-242-8858	Sequence 8858, Ap
32	78	7.3	648	10	US-09-815-242-8986	Sequence 8986, Ap
33	78	7.3	768	9	US-09-938-842A-812	Sequence 812, App
34	78	7.3	1362	10	US-09-784-990-13	Sequence 13, Appli
35	77.5	7.3	2155	10	US-09-960-428-13	Sequence 13, Appli
36	77.5	7.3	3089	10	US-09-880-107-3687	Sequence 3687, Ap
37	77.5	7.3	4509	9	US-10-124-800-5	Sequence 5, Appli
38	77.5	7.3	4512	9	US-10-124-800-27	Sequence 27, Appli
39	77	7.2	735	9	US-09-738-626-1873	Sequence 1873, Ap
40	77	7.2	2169	10	US-09-815-242-6903	Sequence 6903, Ap
41	77	7.2	7499	9	US-09-870-759-105	Sequence 105, App
42	76.5	7.2	483	10	US-09-815-242-3859	Sequence 3859, Ap
43	76.5	7.2	501	10	US-09-815-242-6517	Sequence 6517, Ap
44	76.5	7.2	1317	10	US-09-350-756-5	Sequence 5, Appli
45	76	7.1	924	9	US-09-738-626-137	Sequence 137, App

ALIGNMENTS

RESULT 1
US-10-024-955-6
Sequence 6, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: House Dust Mite and Uses therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996

```

; GENERAL INFORMATION:
; APPLICANT: Cherry, Joel
; APPLICANT: Svendsen, Allan
; APPLICANT: Andersen, Carsten
; APPLICANT: Beler, Lars
; APPLICANT: Frandsen, Torben
; TITLE OF INVENTION: Maltogenic Alpha-Amylase Variants
; FILE REFERENCE: 5443.414-US
; CURRENT APPLICATION NUMBER: US/09/386,607
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: DK98/00269
; EARLIER FILING DATE: 1998-02-27
; EARLIER APPLICATION NUMBER: 60/077,795
; EARLIER FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2160
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2160)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (100)..(2157)
US-09-386-607-1

```

```

Alignment Scores:
Pred. No.:      8.08      Length:      2160
Score:          75.00      Matches:      20
Percent Similarity: 52.46%  Conservative: 12
Best Local Similarity: 32.79%  Mismatches: 21
Query Match:      7.02%      Indels:      8
DB:               4         Gaps:      2

```

US-10-024-955-7 (1-213) x US-09-386-607-1 (1-2160)

```

QY 129 ThHISValIleSerAspIleGlnAspPheValValAlaLeuSerLeuGluIleSerAsp 148
    ||| ::: |||:::||||||| |||||:::
DB 1742 ACTTCAATCGATTGGATGTCACAGATTTCACAGTCGCTGCTACT----- 1698
    |||:::||||:::
QY 149 GluGlyAsnIlePheMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHis 168
    |||:::||||:::
DB 1697 CCGCCAAATGTCACGGTCCCTCGTCGTCGCCAAACCTTCCCGTCGATCGACACACA 1638
    |||:::||||:::
QY 169 IleGlyGlyLeuSerIleLeu-----AspProIlePheGlyValLeuSerAspVal 185
    ::: |||::: ||| ||||| ||| |||:::|||||
DB 1637 TTTACCCGGAATCCCATATTGGAGCAGCACCGATCCGATTGGCGCGCTGAAGCGGATGTG 1578
    |||
QY 186 Leu 186
    |||
DB 1577 CTG 1575

```

Search completed: February 20, 2003, 18:54:08
Job time : 5071 secs

```

; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 114..3149
; US-09-031-655-11

Alignment Scores:
Pred. No.: 12.5      Length: 3230
Score: 75.50        Matches: 36
Percent Similarity: 45.81%  Conservative: 35
Best Local Similarity: 23.23%  Mismatches: 63
Query Match: 7.07%      Indels: 21
DB: 3                Gaps: 9

US-10-024-955-7 (1-213) x US-09-031-655-11 (1-3230)
QY 40 11eGlunSerglUthrIleAspPrometlyValProaPhIslaAspIlyPheglu 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 GTGAGTCGGAGTCTCACAGTAAGTCAAGCACACTCCCTGTGGCGTTTATGCACATAAC 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 ArgHISValGlyIleValaIAspPheIysGlyIleuAla---MetaIrgaSnIIeGluAla 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 477 GGCACCATTAACCGCGTACCTTCAAGGAAGCTGTGAGTGAAGTGAAGTGTAGCTAC 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 ArgGlyLeuIysGlnMetIysArgGlnGlyAspAlaAsnValIysGlyIleGluGlyIle 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 537 AATGGGTTG-----ATGCTGCAACAGCCACATCAACGCAAAATTGGG---581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 ValIysAlaHisLeuLeuIleGlyValaIHisAspAspIleValSerMet-----GluTyr 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 -----AACGTCCTAGTAAGGGAAGGGTTACTGTCTCTACGCTTACCACCATCTAT 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 AspLeuAlaTyr---LysIleuGlyAspLeuHisProThrThrHisValIleSerAspIle 135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 633 GATCTTGGTATGTGAGGCTGTGACCCCATACCCGCTTAGGCGCTT-----680
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 GlAspPheValaIAlaLeuSerLeuGluIleSerAspIleGluIysAnIle---ThrMet 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 681 ----GACCCAAAATGTGAGACACATGTGACAGAGTACAGGCCACAGACTACACCATATA 737
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 ThrSerPheGluValaIArgGlnPheAlaAsnValaIAsnHisIleGlyIleuSerIle 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 738 ACTGACAGCTGATGATTAACAATTCTCATCACAGTAC---CAACAGGTGGGGTAACATC 794
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 LeuAspProIlePheGlyValaLeuSerAspValleuThralaIle 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 795 -----ACCCTGTTCTCAGCCCAACATTGATGATGCATCACAGCGCTC 833
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 49
PCT-US91-03056-16
; Sequence 16, Application PC/TUS9103056
; GENERAL INFORMATION:
; APPLICANT: Vakharla, Vikram
; TITLE OF INVENTION: SPECIFIC DNA AND RNA SEQUENCES
; TITLE OF INVENTION: ASSOCIATED WITH US IDV VARIANTS, VECTOR CARRYING DNA
; TITLE OF INVENTION: SEQUENCES, HOST CARRYING CLONED VECTOR, DEDICED AMINO ACID
; NUMBER OF INVENTION: SEQUENCES, VACCINE AND METHOD OF VACCINATION
; CORRESPONDENCE ADDRESSES: 20
; ADDRESSEE: Viviana Amzel, Ph.D.
; STREET: 112 East Pecan, 2000 NBC Bank Plaza
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: PCT/US91/03056
; FILING DATE: 19910718
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/514,202
; FILING DATE: 14-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel Ph.D., Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: U-0125.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/554-5325
; TELEFAX: 512/226-8395
; TELEX: 762609
; INFORMATION FOR SEQ. ID NO.: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3230 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Infectious bursal disease virus of chickens
; STRAIN: GLS
; IMMEDIATE SOURCE:
; CLONE: GLS-1; GLS-2; GLS-3; GLS-4
; PCT-US91-03056-16

Alignment Scores:
Pred. No.: 12.5      Length: 3230
Score: 75.50        Matches: 36
Percent Similarity: 45.81%  Conservative: 35
Best Local Similarity: 23.23%  Mismatches: 63
Query Match: 7.07%      Indels: 21
DB: 5                Gaps: 9

US-10-024-955-7 (1-213) x PCT-US91-03056-16 (1-3230)
QY 40 11eGlunSerglUthrIleAspPrometlyValProaPhIslaAspIlyPheglu 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 GTGAGTCGGAGTCTCACAGTAAGTCAAGCACACTCCCTGTGGCGTTTATGCACATAAC 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 ArgHISValGlyIleValaIAspPheIysGlyIleuAla---MetaIrgaSnIIeGluAla 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 477 GGCACCATTAACCGCGTACCTTCAAGGAAGCTGTGAGTGAAGTGAAGTGTAGCTAC 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 ArgGlyLeuIysGlnMetIysArgGlnGlyAspAlaAsnValIysGlyIleGluGlyIle 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 537 AATGGGTTG-----ATGCTGCAACAGCCACATCAACGCAAAATTGGG---581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 ValIysAlaHisLeuLeuIleGlyValaIHisAspAspIleValSerMet-----GluTyr 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 -----AACGTCCTAGTAAGGGAAGGGTTACTGTCTCTACGCTTACCACCATCTAT 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 AspLeuAlaTyr---LysIleuGlyAspLeuHisProThrThrHisValIleSerAspIle 135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 633 GATCTTGGTATGTGAGGCTGTGACCCCATACCCGCTTAGGCGCTT-----680
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 GlAspPheValaIAlaLeuSerLeuGluIleSerAspIleGluIysAnIle---ThrMet 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 681 ----GACCCAAAATGTGAGACACATGTGACAGAGTACAGGCCACAGACTACACCATATA 737
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 ThrSerPheGluValaIArgGlnPheAlaAsnValaIAsnHisIleGlyIleuSerIle 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 738 ACTGACAGCTGATGATTAACAATTCTCATCACAGTAC---CAACAGGTGGGGTAACATC 794
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 LeuAspProIlePheGlyValaLeuSerAspValleuThralaIle 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 795 -----ACCCTGTTCTCAGCCCAACATTGATGATGCATCACAGCGCTC 833
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 50
US-09-386-607-1/c
; Sequence 1, Application US/09386607
; Patent No. 6162628

```

```

Db 417 GTGAGTGGAGATCTCACAGTAAGGTAACGACAACTCCCTGGTGGCTTTATGACATAAC 476
Oy 60 ArgHisValGlyIleValAspPheIysGlyIleValA---MetArgsnIleGluA 78
Db 477 GGCACCAATAAAGCCGACCTTCCAAAGAAAGCCTGAGTAACATGACAGATGTTACCTAC 536
Oy 79 ArgGlyLeuIysGlnMetIysArgGlnGlyAspAlaAsnValLysGlyGlnGlyIle 98
Db 537 AATGGCTTG-----ATGCTGCAACAGCCAACTCAACGACAATAATTGGG---581
Oy 99 ValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMet-----GluTyr 116
Db 582 -----AACGTCCTAAGTAGGGAAGGGGTTACTGTCTCTCAAGCTTACCAACATCATAT 632
Oy 117 AspLeuAlaTyr---LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIle 135
Db 633 GATCTTGGAATGTGAGAGCTTGGTGGACCCCATACCCCTATAGAGGCT-----680
Oy 136 GlnAspPheValValAlaLeuSerLeuGlnIleSerAspIleGlyAsnIle---ThrMet 154
Db 681 ---GACCCAAAATGATGACAACATGTGACAGCATGACAGGCCCAAGATCTACACCATTA 737
Oy 155 ThrSerPheGluValArgGlnPheAlaAsnValValHisHisIleGlyIleSerIle 174
Db 738 ACTGCACTGATGATTACCAATTCTCTACACAGAC---CAACAGGTGGGGTAACATC 794
Oy 175 LeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIle 189
Db 795 -----ACCTGTCTCAGCCAACTGATGATCATCATACAGCCTC 833

RESULT 48
US-09-031-655-11
: Sequence 11, Application US/09031655
: Patent No. 6017759
: GENERAL INFORMATION:
: APPLICANT: VAKHARIA, VIKRAM
: APPLICANT: SNYDER, DAVID B
: APPLICANT: MENDEL-WHERSAT, STEPHANIE A
: TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
: TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
: TITLE OF INVENTION: THEROON
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLOH, SPIYAK, MCCLELLAND, MAIER & NEUSTADT
: STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
: CITY: ARLINGTON
: STATE: VIRGINIA
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/031,655
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/219,262
: FILING DATE: 29-MAR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: OBLOH, NORMAN F
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 2747-047-27
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 413-3000
: TELEFAX: (703) 413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3230 base pairs
: type: nucleic acid

```


DB: 4 Gaps: 7

US-10-024-955-7 (1-213) x US-08-934-254-1 (1-3588)

QY 33 ILeAspAspAlaIleAlaAlaIleGluInSerGluThrIleAspPromElytysValPro 52
:::|||||
:::|||||

Db 743 CTGAATGATTTCAATCCCTTGCGAGTCGCTTAAGACAGTTTTTGGATGCGGCGAAGTTACCC 802

QY 53 AsphIsAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyIleuAla 72
|||
|||

Db 803 GAT-----CGCATCACATCATCTATTTGTGGCGCTGGGGGAGTGAAGC 844

QY 73 MetArgAsnIleGlu----- 77
|||
|||

Db 845 ATGGCCATTATTGAAGATTAAATTCACCAGGCCCATGAATAATGTGTAATCGAAAAAGAT 904

QY 78 -----AlaArgGlyLeu----- 81
|||
|||

Db 905 ACAGATTAATCGTTCTTCATACGCGCCGCTCCCTGGGGGTGCCCGTAATTGTGAGAGAT 964

QY 82 ---LysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys 100
::: |||
::: |||

Db 965 GCCCGCTAGAAAGAAAGCTTGGCTGCGCCAAATATCAACGACCGAAGCCATATGTG--- 1021

QY 101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetLeuLysPheAlaIleArg 120
::: |||
::: |||

Db 1022 -----GtGGCCACCAGCAGCAGACCGCTTAACCTTGAAATTTGGCTTAACCTGCC 1069

QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAspIleGlnAspPheVal 140
|||
|||

Db 1070 AAGCG-----ATCGCCCTAGCGCTCCAGTGTGTGGCTTGGCCAGATGCCAGTTT 1123

QY 141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArg 160
::: |||
::: |||

Db 1124 AGCCTGTCCTCGCAGAAAGATTGTAATTTGAAGAAGGTGCTTTGTCGGCGGAATTGGCC 1183

QY 161 GlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
|||
|||

Db 1184 ACCTATTCTTTTGGCGGCGCGCCCTGGGGGC--AAATTTTGGGCAAC-----GGC 1234

QY 181 ValLeuSerAspValLeu 186
::: |||
::: |||

Db 1235 ATGACCGATGATTGTCTG 1252

RESULT 41
US-08-961-527-97/c
: Sequence 97, Application US/08961527
: Patent No. 6420135

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

```

:      REGISTRATION NUMBER: 36,373
:      REFERENCE/DOCKET NUMBER: PB34091
:      TELECOMMUNICATION INFORMATION:
:          TELEPHONE: (301) 309-8504
:          TELEFAX: (301) 309-8512
:      INFORMATION FOR SEQ ID NO: 97:
:      SEQUENCE CHARACTERISTICS:
:          LENGTH: 9069 base pairs
:          TYPE: nucleic acid
:          STRANDEDNESS: double
:          TOPOLOGY: linear
:      US-08-961-527-97
:
Alignment Scores:
Pred. No.:      47.5        Length:    9069
Score:         76.00       Matches:     58
Percent Similarity:   41.56%   Conservative: 38
Best Local Similarity: 25.11%   Mismatches: 81
Query Match:      7.12%       Indels:    54
DB:              Gaps:      14
US-10-024-955-7 (1-213) x US-08-961-527-97 (1-9069)
QY      1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
      :::::::::::::::::::: ||| |||
Db      2387 GTTATGAAAGTAATTGTTC-----GCATAATGGTTTC-----CTT 2355S
QY      21 HistyrAspLysIleThr-----GIUGLIUleasnLySaLaileAspAlaIle 37
      ||| |:::||||| |::| :|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      2354 CACGCAAGATCGCATCACAATCTACTGGTAGACGTAGCTGAAGAASCTGGCTGACTTGCA 2295S
QY      38 AlaAlaIleGlucInserGIuThrIIeasPrrometLysValProAspHisAlaSprly 57
      ::||| |::|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      2294 GACTTGCACCTCACGCCAAMAAATTAATCAATGACCCTTGAAAATCCAAACSTGCCGATGT 2235S
QY      58 PheGluArgHisValGIuIleValAspPheLysGIuLuleAlaMetArgsnIIegLu 77
Db      2234 -----CCGCTTATCATCTTGAAACGGSAACSTTGCTGCTGACSGTCACTT 2190S
QY      78 AlaArg-----GIyleuLysImetLysAtgIngIyAspAlaAsnVal---LysGIu 94
      |||::|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      2189 GCCAAGATACAGGTGTAAAGTGGCGTCGACAGCTGGGSCAACGTAAGGTGTTTGACTCA 2130S
QY      95 GluGIuGIuIleValLySaLanIleSuLeuIleGIuValnHIsAsp---AspIleValSer 113
      ||||||| ::::|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      2129 GAAGAAGATCGCATCTACGCCCTTCTGACAGATGAAGTAATGTTTGAATGCCATGACTGCTT 2070T
QY      114 MetcIUuArgspReuIlArGuILySLeuGIuYaRLeuHIlsProthIHnisValIleSer 133
Db      2069 GTTTCGTTTT---GTGGACSTTAAGGTGGTGGCTGGTATGCTCTGAC----- 2028S
QY      134 AspIleGlnAspPheValAlaIleUsenSeuGIuIleSerAspJcuGLyAsn----- 151
Db      2027 -----ATGCAATACATCTTTCTCATATGATGATTGTTGGTAAGAAGCTCAGGAGAT 1983S
QY      152 -----lIethrMetTriserPheGIuValArgIlnPheHlaSnValyAlsn 167
Db      1982 AAGGTGCCCCCTTGTACGGACGGACGACCTTCTGTGTGTACTTAAGTCTGCTGTGTGA 1923S
QY      168 HisIle-----GIugLYleuSerIIleuIdasProIlePheGIuVal 181
      ||||||| |::|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      1922 CATATTCCTCTCGAAGCTCAGATGGTGG-----CCAATTCCTATCTC 1878S
QY      182 LeuserpValIleuthraIleIepheGlnaspThValArgLySGImetHTrgVal 201
      ::|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      1877 CGTACCGCGCATATCGTTACGGTTGACCAAAGATAC-----AAGAAATTTCTATGGCC 1822S
QY      202 LeuAlaProAlaPheLysArgGIuLeuGIuLys 212
      ::::|:~::~||| |||||||
Db      1823 GTATCC-----GAAGAAGAACTTTGA AAAA 1800

```

```

APPLICANT: Nuccio, Michael
APPLICANT: Freysinet, Georges L.
APPLICANT: Nunberg, Andrew N.
TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
TITLE OF INVENTION: DELTA 6-DESATURASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,936
FILING DATE: 28-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/366,779
FILING DATE: 30-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Presser, Leopold
REGISTRATION NUMBER: 19,827
REFERENCE/DOCKET NUMBER: 83832YXW
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3588 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2002..3081
US-08-789-936-1

Alignment Scores:
Pred. NO.: 12.5 Length: 3588
Score: 76.00 Matches: 44
Percent Similarity: 37.63% Conservative: 26
Best Local Similarity: 23.66% Mismatches: 68
Query Match: 7.12% Indels: 48
DB: 1 Gaps: 7

US-10-024-955-7 (1-213) x US-08-789-936-1 (1-3588)
QY 33 llaaspalalailealaiailegluglinsergluThrllaapPrometLysValPro 52
Db ::::||||| ||||| ::::: ||||| ||||| |||||
743 CTGATGATTCACCTTGGCGAGTGCGCTTAGTCAGTTTGTGGATGCGGCCAAGTTVACC 8020
QY 53 AsphsAlaaaprlYsphegluarGHisValgLytleValasprhelYsglygluleAla 72
Db ||||| ||||| ||||| ||||| ||||| ||||| |||||
803 GAT-----CGCATCACATCAATCTTTGTGGCGCTGGGGGAGTGAGC 844
QY 73 Metarganilegu-----Alaargyleu----- 77
Db ||| ||||| ||||| ||||| ||||| ||||| |||||
845 ATGGCATTAATTGAAGAATAATTCACCAAGGCCATGAATTTGTAATCGAAAAGAT 904
QY 78 -----Alaargyleu----- 81
Db 905 ACAGATTAATCGTTTCTTCATACAGCCGCGCTCCTGGGGGATGCCGTAATTGTGAGAT 964
QY 82 ---LysginMetLysArgGlnGlyAspAlaIasnValylSgLYslnglulglylleVallys 100
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db      965 GCCGCCCTAGAAAGAACGTTGGCCCTCGCCCAATATCAACGCGAGCCGAAGCCATTGTG--- 1021
Oy      101 AlahisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaIArg 120
Db      1022 -----GtGGCCACACGACGACACCGCTTAACTTGGAATTTGGCAATTTGGCTACTCTGCC 1063
Oy      121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValVal 140
Db      1070 AAGCG-----ATCGCCCTACCTGCCAGTGTGTGTTCGCTTGCAGAGATCCCAAGTT 1122
Oy      141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArg 160
Db      1124 AGCCTGTCCTCCGACGAAAGATTTGAATTTGAACCGCTGCTTTGTCGCGGGAATTTGGCC 1183
Oy      161 GlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
Db      1184 ACCTATTCTCTTGGCGGCGGCGCCCTGGGGGCG--AAATTTTGGGCAAC-----GCG 1234
Oy      181 ValLeuSerAspValLeu 186
Db      1235 ATGACCGATGATTTCGTG 1252

RESULT 40
US-08-934-254-1
: Sequence 1, Application US/08934254
: Patent No. 6355861
:
: GENERAL INFORMATION:
: APPLICANT: Thomas, Terry L.
: TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
: TITLE OF INVENTION: DELTA 6-DESATURASE
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Scully, Scott, Murphy & Presser
: STREET: 400 Garden City Plaza
: CITY: Garden City
: STATE: New York
: COUNTRY: United States
: ZIP: 11530
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/934,254
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Presser, Leopold
: REGISTRATION NUMBER: 19,827
: REFERENCE/DOCKET NUMBER: 8383ZYXWVU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (516) 742-4343
: TELEFAX: (516) 742-4366
: TELEX: 230 901 SANS UR
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3588 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2002..3081
: US-08-934-254-1

Alignment Scores:
Pred. No.: 12.5 Length: 3588
Score: 76.00 Matches: 44
Percent Similarity: 37.63% Conservative: 26
Best Local Similarity: 23.63% Mismatches: 68
Query Match: 7.12% Indels: 48

```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,779
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Presser, Leopold
; REGISTRATION NUMBER: 19,827
; REFERENCE/DOCKET NUMBER: 83832YXW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2002..3081
; US-08-366-779-1

Alignment Scores:
Score: 12.5 Length: 3588
Percent Similarity: 76.00 Matches: 44
Best Local Similarity: 37.63% Conservative: 26
Query Match: 23.66% Mismatches: 68
DB: 7.12% Indels: 48
Gaps: 7

US-10-024-955-7 (1-213) x US-08-366-779-1 (1-3588)
QY 33 IleASPAlaIleAlaIleIleGluInserGluThrIleASPProMetIysValPro 52
DB 743 CTGAATGATTTCATCTTGGCAGTCCGCTTATGTCAGTTTGGATCGGCCAAGTTACCC 802
QY 53 ASPHISAlaAspIysPheGluArgHisValIleGlyIleValASPheIysGluIleVala 72
DB 803 GAT-----CGCCATCATCATCATTTGGGCTGGGGGAGAGGAGC 844
QY 73 MetArgAsnIleGlu----- 77
DB 845 ATGGCCATTATTGAAGAGTTAATTCACAGGCCCATGAATTGTGCTAATCGAAGAGAT 904
QY 78 -----AlaArgGlyLeu----- 81
DB 905 ACAGATATATCGTTTCTTCGATACGCGCCCTCCCTGGGGTCCCTAATTGTGAGAGAT 964
QY 82 ---LysGluMetIysArgGlnGlyAspAlaAsnValIysGlyGluIleValIys 100
DB 965 GCCCGCTAGAAAGACGTGGCGCTCGCCCAATATCAACGCGAAGCCAGCATTTGG--- 1021
QY 101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetCysIuTrAspLeuAlaTyr 120
DB 1022 -----GTGGCCACACGAGCGACACCGCTTAATCTTGAATTGGCTAATGCTGCC 1069
QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAspIleGlnAspPheValVal 140
DB 1070 AAGGGC-----ATGCCCCCTAGCCTCGCAGTGGTGTTCGTTGCGAGATGCCAGTTT 1123
QY 141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArg 160
DB 1124 ACCCTGTCCCTCAGAGAAATATTGAATTGAACGCTCTTTGCGCGGCAATTTGCC 1183
QY 161 GluPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
DB :----- 11

```

```

DB 1184 ACCTATTCCTTTCGGCGGCGCCCTGCGGGGC---AAATTTTGGGCAAC-----GGC 1234
QY 181 ValLeuSerAspValLeu 186
DB 1235 ATGACCGATGATTTGCTG 1252

RESULT 37
US-08-478-727-1
; Sequence 1, Application US/08478727
; Patent No. 5653068
; GENERAL INFORMATION:
; APPLICANT: Thomas, Terry L.
; APPLICANT: Reddy, Avutu S.
; APPLICANT: Nuccio, Michael
; APPLICANT: Freysinet, Georges L.
; TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC
; TITLE OF INVENTION: ACID BY A DELTA 6-DESATURASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,727
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/307,382
; FILING DATE: 14-SEP-1994
; APPLICATION NUMBER: US 07/817,919
; FILING DATE: 08-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McNulty, William E.
; REGISTRATION NUMBER: 22,606
; REFERENCE/DOCKET NUMBER: 83832
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2002..3081
; US-08-478-727-1

Alignment Scores:
Score: 12.5 Length: 3588
Percent Similarity: 76.00 Matches: 44
Best Local Similarity: 37.63% Conservative: 26
Query Match: 23.66% Mismatches: 68
DB: 7.12% Indels: 48
Gaps: 7

US-10-024-955-7 (1-213) x US-08-478-727-1 (1-3588)
QY 33 IleASPAlaIleAlaIleIleGluInserGluThrIleASPProMetIysValPro 52
DB 743 CTGAATGATTTCATCTTGGCAGTCCGCTTATGTCAGTTTGGATCGGCCAAGTTACCC 802

```

```

QY 125 LeuHisProThrThrHisValIleSerAspIleGlnAspPheValAlaIleSerLeu 144
Db 1172 -----CATGGTGAAGTCCGTA---GAACCTGTAGTAAGAGAGTCCAT 1131
QY 145 GluIleSerAspGluGlnIleThrMetThrSerPheGluValArgGlnPheAlaAsn 164
Db 1130 AGAGCACTGCATCAAGTTACCCAGAGAGCTTCACCTGCTGACCAATTC----- 1077
QY 165 ValValAsnHisIleGlyGlyLeuSer-----IleLeuAsp 176
Db 1076 -----AACTTCTGGCCAGATTAGTGTAGACGTAGCCAGATCCAGATCGATCGCGGAT 1023
QY 177 ProIlePhe-----GlyValLeuSerAspValLeuThrAlaIlePheGln 191
Db 1022 CCGCTTTTGTAGTCTTTAACTAGGTAGTACGTAGACATCTGTGTGCGCGTACT 963
QY 192 AspThrValArgGlyGlyLeuThr----- 199
Db 962 CTGACCATAGGTAGAAATAGTCTCCAAATTAGCGAAGTTGGCGCTGGTGGGAATG 903
QY 200 -----LysValIleuAlaProAlaPheLysArg 208
Db 902 GAGACCGTTGGGTAGACAGAGATGGGTAGTACCCAGTTCCAAAG 855

RESULT 35
US-08-307-382-1
; Sequence 1, Application US/08307382
; Patent No. 5552306
; GENERAL INFORMATION:
; APPLICANT: Thomas, Terry L.
; APPLICANT: Reddy, Avutu S.
; APPLICANT: Nuccio, Michael
; APPLICANT: Freysinet, Georges L.
; TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC
; TITLE OF INVENTION: ACID BY A DELTA 6-DESATURASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,382
; FILING DATE: 14-SEP-1994
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/817,919
; FILING DATE: 08-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McNulty, William F.
; REGISTRATION NUMBER: 22,606
; REFERENCE/DOCKET NUMBER: 8383Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4366
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: 2002...3081
US-08-307-382-1
Alignment Scores:
Score: No.: 12.5 Length: 3588
Score: 76.00 Matches: 44
Percent Similarity: 37.63% Conservative: 26
Best Local Similarity: 23.66% Mismatches: 68
Query Match: 7.12% Indels: 48
Db: 1 Gaps: 7

US-10-024-955-7 (1-213) x US-08-307-382-1 (1-3588)
QY 33 IleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspPheValPro 52
Db 743 CTGAATGATTTTCATCCCTTGCGACATCGCTTGTAGACATTTTGGATCGGCGCAAGTACCC 802
QY 53 AspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyIleuAla 72
Db 803 GAT-----CCGCATCATCATCATCATTTGTGGCTGGGGGAGTGGAGC 844
QY 73 MetArgAsnIleGlu-----AlaArgGlyLeu----- 77
Db 845 ATGCGCATTTATTTGAAGATTAAATTACACAGGCCATGAATGTTGGTAATCGAAAGAT 904
QY 78 -----AlaArgGlyLeu----- 81
Db 905 ACAGATAATCGTTTCTTGATACGAGCGCCGCTCCCTGGGGGTGCCGTAATTGGAGAGAT 964
QY 82 ---LysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyIleValLys 100
Db 965 GCCCGCTGAAGAAGAGTGGCGCTGCCCAATATCACCGACCGCAAGCATTTG--- 1021
QY 101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGlnArgLysPheuAlaTyr 120
Db 1022 -----GTGCCACACGACGACCGCTTAACTTGAATTTGGCTTAATGCTGCC 1069
QY 121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValAla 140
Db 1070 AAGGCG-----ATGCCCTTACCGTCGCGCAGTGTGTGGTGGCCAGATGCCAGTTT 1123
QY 141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArg 160
Db 1124 ACCCTGCTCCCTGCAGAGATTTGAATTTGAACGCTGCTTGTCCGGCGAATTGGCC 1183
QY 161 GlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
Db 1184 ACCTATTCCTTTGGCGCGCGCGCCGTGGGGGC---AAATTTTGGGCAAC-----GGC 1234
QY 181 ValLeuSerAspValLeu 186
Db 1235 ATGACCGATGATTTGCTG 1252

RESULT 36
US-08-366-779-1
; Sequence 1, Application US/08366779
; Patent No. 5614393
; GENERAL INFORMATION:
; APPLICANT: Thomas, Terry L.
; APPLICANT: Reddy, Avutu S.
; APPLICANT: Nuccio, Michael
; APPLICANT: Freysinet, Georges L.
; APPLICANT: Nunberg, Andrew N.
; TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
; TITLE OF INVENTION: DELTA 6-DESATURASE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530

```


MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/708,541A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: KITTS, Monica C.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P8172-6002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/638-5000
TELEFAX: 202/638-4810
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3261 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 97..531
US-08-708-541A-27

Alignment Scores:
Pred. No.: 8.15 Length: 3261
Score: 77.00 Matches: 39
Percent Similarity: 44.91% Conservative: 36
Best Local Similarity: 23.35% Mismatches: 64
Query Match: 7.21% Indels: 28
DB: 2 Gaps: 10

US-10-024-955-7 (1-213) x US-08-708-541A-27 (1-3261)

QY 40 IlegluInserclutrhrlleasprPrometlyValProasphIsalaasplyspheglu 59
DB 434 GTGAGTCGGAGTCACAGTCAAGACACTTCCTGCGCTTATGACACTAAC 493
QY 60 ArgHsValGlyLeuValAspHelysGlyLeuAla--MetArgasnIlegluAla 78
DB 494 GGCACCATTAAGCCCTGACCTTCAAGAAAGCCTGAGTGAAGTACAGATGTACCTAC 553
QY 79 ArgGlyLeuValAspHelysArgGlnGlyAspAlaAsnValysGlyGluGluGly 98
DB 554 AATGGGTG-----ATGCTGCAACAGCCCAACATCAACGACAAATTTGG--- 598
QY 99 VallyAlaHisLeuLeuIleglyValHisAspAspIleValSerMet-----GluTyr 116
DB 599 -----AACGTCTCTAGTAGGGAAGGGGTCCAGCTCTCAGCTTACCCACATCATAT 649
QY 117 AspLeuAlaTyr--LysLeuGlyAspLeuHisProThrThriValIleSerAspIle 135
DB 650 GATCTGGGTATGTGAGCTGTGTGACCCCATTCGCGCAATAGAGCTT----- 697
QY 136 GlnAspPheValAlaLeuSerLeuGluIleSerAspGluLysIle---ThrMet 154
DB 698 ---GACCCAAATAATGGTAAACACATGTGACAGCAGGAGCCGACAGATCTACACCA 754
QY 155 ThrSerPheGluValArgGlnPheAlaAsnValAlaHisIleIleglyLeuSerIle 174
DB 755 ACTGCAGCGCATATTCACATATTCATCATCAGTACCAACCA---GTTGGGATTAACATC 811
QY 175 LeuAspProIlePheGlyValLeuSerAspValLeuThrAla----- 188
DB 812 -----ACACTGTTCTCAGCCCAACATTAATGATGTCATACAAAGCTCAGCGTTGGGAGAG 865
QY 189 ---IlePheGlnAspThrVal 194
DB 866 CTCGTGTTCAACAAAGCGTC 886

RESULT 32

US-08-708-541A-29

Sequence 29, Application US/08708541A

Patent No. 5871744

GENERAL INFORMATION:

APPLICANT: VAKHARIA, Vikram N.

APPLICANT: MUNDT, Egbert

TITLE OF INVENTION: A METHOD FOR GENERATING BIRNAVIRUS FROM

TITLE OF INVENTION: SYNTHETIC RNA TRANSCRIPTS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIKALDO, MARCELSTEIN, MURRAY & ORAM LLP

STREET: 655 Fifteenth Street, N. W.,

CITY: Suite 330 - G Street Lobby

STATE: DC

COUNTRY: USA

ZIP: 20005-5701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/708,541A

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: KITTS, Monica C.

REGISTRATION NUMBER: 36,105

REFERENCE/DOCKET NUMBER: P8172-6002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202/638-5000

TELEFAX: 202/638-4810

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 3261 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 131..3166

US-08-708-541A-29

Alignment Scores:
Pred. No.: 8.15 Length: 3261
Score: 77.00 Matches: 39
Percent Similarity: 44.91% Conservative: 36
Best Local Similarity: 23.35% Mismatches: 64
Query Match: 7.21% Indels: 28
DB: 2 Gaps: 10

US-10-024-955-7 (1-213) x US-08-708-541A-29 (1-3261)

QY 40 IlegluInserclutrhrlleasprPrometlyValProasphIsalaasplyspheglu 59
DB 434 GTGAGTCGGAGTCACAGTCAAGACACTTCCTGCGCTTATGACACTAAC 493
QY 60 ArgHsValGlyLeuValAspHelysGlyLeuAla--MetArgasnIlegluAla 78
DB 494 GGCACCATTAAGCCCTGACCTTCAAGAAAGCCTGAGTGAAGTACAGATGTACCTAC 553
QY 79 ArgGlyLeuValAspHelysArgGlnGlyAspAlaAsnValysGlyGluGluGly 98
DB 554 AATGGGTG-----ATGCTGCAACAGCCCAACATCAACGACAAATTTGG--- 598
QY 99 VallyAlaHisLeuLeuIleglyValHisAspAspIleValSerMet-----GluTyr 116
DB 599 -----AACGTCTCTAGTAGGGAAGGGGTCCAGCTCTCAGCTTACCCACATCATAT 649
QY 117 AspLeuAlaTyr--LysLeuGlyAspLeuHisProThrThriValIleSerAspIle 135

```
QY 151 AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGly 170
      ::::: 111 111
DB 205080 TTCCTTACCGGACCGCATTCCTCC-----GGT 205106

QY 171 GlyLeuSerIleLeuAspProIlePheGly 180
      111 111: 111111
DB 205107 GCGTTCGCGCCCTCGACAGCATCTTTGGC 205136

RESULT 29
US-09-425-453A-7
; Sequence 7, Application US/09425453A
; Patent No. 6468793
; GENERAL INFORMATION:
; APPLICANT: Teem, John L.
; TITLE OF INVENTION: CFTF Genes and Proteins for Cystic Fibrosis Gene Therapy
; FILE REFERENCE: FSU-99XC1
; CURRENT APPLICATION NUMBER: US/09/425,453A
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/105,444
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 4443
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1)..(4443)
US-09-425-453A-7

Alignment Scores:
Pred. No.: 11 Length: 4443
Score: 77.50 Matches: 35
Percent Similarity: 45.53% Conservative: 21
Best Local Similarity: 28.46% Mismatches: 50
Query Match: 7.26% Indels: 17
DB: 4 Gaps: 7

US-10-024-955-7 (1-213) x US-09-425-453A-7 (1-4443)
QY 97 GlyIleValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyr 116
      ::::: 111 111: 11111
DB 1498 GGCACCATTTAAAGAAATATCATCTTTGGTGT-----TCCTAT 1536

QY 117 AspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGln 136
      111 111: 111: 11111
DB 1537 GAT---GAATATATGATATACAGAACGTCATCAAGCATGCCAAGTACAGAGCATCTCC 1593

QY 137 AspPheValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSer 156
      111 111: 111: 11111
DB 1594 AAGTTTGCAGAGAAAGACATATG---GTTCTTGGAGAGAGTGGAATCACACTGAGTGA 1650

QY 157 PheGluValArgGln-----PheAlaAsnValValAsnHisIleGlyLeuSerIle 174
      ::::: 111 111 111
DB 1651 GGTCAATATGCAAGATTTCTTTAGCAAGAGCATATCAAGATCTGATTTGATTTA 1710

QY 175 LeuAspProIlePheGlyValLeuSerAspValLeuThr-----AlaIlePheGlnAsp 192
      111111 111111 111 111111111 111111:
DB 1711 TTAGACTCTCTTTGGATACCTA---GATGTTTAAACAGAAAGAAATATTATTTGAAGC 1767

QY 193 ThrValArgLysGluMet-----ThrLysValLeuAlaProAlaPheLysArgGlu 209
      111 111 111 111: 111: 111
DB 1768 TGTGCTGTAAAGTGAAGCTTAACAAAGTATTTGCTACTCTTAAATGGAACAT 1827

QY 210 LeuGluLys 212
      111: 111: 111
DB 1828 TTAAAGAAA 1836

RESULT 30
US-09-425-453A-17
; Sequence 17, Application US/09425453A
; Patent No. 6468793

; GENERAL INFORMATION:
; APPLICANT: Teem, John L.
; TITLE OF INVENTION: CFTF Genes and Proteins for Cystic Fibrosis Gene Therapy
; FILE REFERENCE: FSU-99XC1
; CURRENT APPLICATION NUMBER: US/09/425,453A
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/105,444
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 4443
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-425-453A-17

Alignment Scores:
Pred. No.: 11 Length: 4443
Score: 77.50 Matches: 35
Percent Similarity: 45.53% Conservative: 21
Best Local Similarity: 28.46% Mismatches: 50
Query Match: 7.26% Indels: 17
DB: 4 Gaps: 7

US-10-024-955-7 (1-213) x US-09-425-453A-17 (1-4443)
QY 97 GlyIleValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyr 116
      ::::: 111 111: 11111
DB 1498 GGCACCATTTAAAGAAATATCATCTTTGGTGT-----TCCTAT 1536

QY 117 AspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGln 136
      111 111: 111: 11111
DB 1537 GAT---GAATATATGATATACAGAACGTCATCAAGCATGCCAAGTACAGAGCATCTCC 1593

QY 137 AspPheValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSer 156
      111 111: 111: 11111
DB 1594 AAGTTTGCAGAGAAAGACATATG---GTTCTTGGAGAGAGTGGAATCACACTGAGTGA 1650

QY 157 PheGluValArgGln-----PheAlaAsnValValAsnHisIleGlyLeuSerIle 174
      ::::: 111 111 111
DB 1651 GGTCAATATGCAAGATTTCTTTAGCAAGAGCATATCAAGATCTGATTTGATTTA 1710

QY 175 LeuAspProIlePheGlyValLeuSerAspValLeuThr-----AlaIlePheGlnAsp 192
      111111 111111 111 111111111 111111:
DB 1711 TTAGACTCTCTTTGGATACCTA---GATGTTTAAACAGAAAGAAATATTATTTGAAGC 1767

QY 193 ThrValArgLysGluMet-----ThrLysValLeuAlaProAlaPheLysArgGlu 209
      111 111 111 111: 111: 111
DB 1768 TGTGCTGTAAAGTGAAGCTTAACAAAGTATTTGCTACTCTTAAATGGAACAT 1827

QY 210 LeuGluLys 212
      111: 111: 111
DB 1828 TTAAAGAAA 1836

RESULT 31
US-08-708-541A-27
; Sequence 27, Application US/08708541A
; Patent No. 5871744
; GENERAL INFORMATION:
; APPLICANT: VAKHARIA, Vikram N.
; APPLICANT: MUNDT, Egbert
; TITLE OF INVENTION: A METHOD FOR GENERATING BIRNAVIRUS FROM
; TITLE OF INVENTION: SYNTHETIC RNA TRANSCRIPTS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM LLP
; STREET: 655 fifteenth street, N. W.,
; STREET: Suite 330 - G Street Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
```



```

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

Alignment Scores:
Pred. No.: 1.74e+05 Length: 4403765
Score: 78.00 Matches: 50
Percent Similarity: 34.76% Conservative: 23
Best Local Similarity: 23.81% Mismatches: 67
Query Match: 7.30% Indels: 70
DB: 4 Gaps: 10

```

```
US-10-024-955-7 (1-213) x US-09-103-840A-2 (1-4403765)
```

```

QY 11 Alphevalalavalserlasprrllehis-----Tyraaplystle 25
Db 204764 GCCATCGTACAGGGGGGGGGGAGATCCAGATCCGGGCTTTCTGGCAAGCTGCACACCTTC 204823
QY 26 ThrGluGluLeuSlnsAlaIleAspAspAlaIleAlaIleGlu----- 41
Db 204824 ACCGACGACCTACACGACGACGACGATGACATTACCGCGGCATTATTCACCAATCGG 204883
QY 42 -----GlnserGluThrIleAsp-----Pro 48
Db 204884 TTGTGGCTTATGTGGGCGGTGCTGGAGATCTCAATCGGCTGCACCGACTACCG 204943
QY 49 MetLysValProAspRhisAlaAspLysPheGluArgHisValGlyIleValAspRheLys 68
Db 204944 CCATTGATACACACTTTCCGGATAGCAGAACTGTGATCAAGCTTCGCGGTG 205003
QY 69 GlyGluLeuAlaMetArgAsnIleGlu-----AlaArgGly----- 80
Db 205004 GCGCGGCTACGACAGTCCGCGACAGATCTTTGGTCCGCGGCGGCGATGTCACACG 205063
QY 81 -----LeuLysGlnMetLysArgGlnGlyAspAlaAsnVal 92
Db 205064 GACCTGCAGGCGCTGCATATGCCGCTCAAGAACTGCTGCA----- 205105
QY 93 LysGluGluGluGlyIleValLysAlaHisLeuLeuIleGlyValHisAspRheLys 112
Db 205106 -----GCCGCTCCGATCTGTGGTGGGCGCTCAAAATTGATCTTC 205144
QY 113 SerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIle 132
Db 205145 ACCGACCCCTTTGACGCTC-----GACACCGCTGCGGACGCTG---GTGGCG 205186
QY 133 SerAspIleGlnAspPheValValAlaLeuSerLeuGluIleSer-----AspGluGly 150
Db 205187 GCGGACTACATGAATCTTGCTGACGCTGACCTGACCTACAGCGCCATGCAATGCG 205246
QY 151 AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGly 170
Db 205247 TTCCTTACCGGACCGGATTCCTC-----GGT 205273

```

```

QY 171 GlyLeuSerIleLeuAspProIlePheGly 180
Db 205274 GCGTTCGCGGCGCTGCAGCAGTCTTTGGC 205303

```

```
RESULT 28
```

```
US-09-103-840A-1
; Sequence 1, Application US/09103840A
```

```

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Alignment Scores:
Pred. No.: 1.74e+05 Length: 4411529
Score: 78.00 Matches: 50
Percent Similarity: 34.76% Conservative: 23
Best Local Similarity: 23.81% Mismatches: 67
Query Match: 7.30% Indels: 70
DB: 4 Gaps: 10

```

```
US-10-024-955-7 (1-213) x US-09-103-840A-1 (1-4411529)
```

```

QY 11 Alphevalalavalserlasprrllehis-----Tyraaplystle 25
Db 204597 GCCATCGTACAGGGGGGGGAGATCCAGATCCGGGCTTTCTGGCAAGCTGCACACCTTC 204656
QY 26 ThrGluGluLeuSlnsAlaIleAspAspAlaIleAlaIleGlu----- 41
Db 204657 ACCGACGACCTACACGACGACGACGATGACATTACCGCGGCATTATTCACCAATCGG 204716
QY 42 -----GlnserGluThrIleAsp-----Pro 48
Db 204717 TTGTGGCTTATGTGGGCGGTGCTGGAGATCTCAATCGGCTGCACCGACTACCG 204776
QY 49 MetLysValProAspRhisAlaAspLysPheGluArgHisValGlyIleValAspRheLys 68
Db 204777 CCATTGATACACACTTTCCGGATAGCAGAACTGTGATCAAGCTTCGCGGTG 204836
QY 69 GlyGluLeuAlaMetArgAsnIleGlu-----AlaArgGly----- 80
Db 204837 GCGCGGCTACGACAGTCCGCGACAGATCTTTGGTCCGCGGCGGCGATGTCACACG 204896
QY 81 -----LeuLysGlnMetLysArgGlnGlyAspAlaAsnVal 92
Db 204897 GACCTGCAGGCGCTGCATATGCCGCTCAAGAACTGCTGCA----- 204938
QY 93 LysGluGluGluGlyIleValLysAlaHisLeuLeuIleGlyValHisAspRheLys 112
Db 204939 -----GCCGCTCCGATCTGTGGTGGGCGCTCAAAATTGATCTTC 204977
QY 113 SerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIle 132
Db 204978 ACCGACCCCTTTGACGCTC-----GACACCGCTGCGGACGCTG---GTGGCG 205019
QY 133 SerAspIleGlnAspPheValValAlaLeuSerLeuGluIleSer-----AspGluGly 150
Db 205020 GCGGACTACATGAATCTTGCTGACGCTGACCTGACCTACAGCGCCATGCAATGCG 205079

```

```
Db 29367 AAAGCTATGTTAGCGCTAGCTAAAGGCTTATTCCTGCA 29405
||| ||| ||| |||||
US-09-232-479-13
; Sequence 13, Application US/09232479
; Patent No. 6221362
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, JEAN-CHRISTOPHE
; APPLICANT: BOUCHARON, ANNABELLE
; APPLICANT: RIVIERE, MICHEL
; TITLE OF INVENTION: AVIAN POLYNUCLEOTIDE VACCINE FORMULA
; FILE REFERENCE: 454313-2260
; CURRENT APPLICATION NUMBER: US/09/232,479
; PRIOR FILING DATE: 1999-01-15
; EARLIER APPLICATION NUMBER: 96/09339
; EARLIER FILING DATE: 1996-07-19
; EARLIER APPLICATION NUMBER: PCT/FR97/01326
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Infectious bursal disease virus
US-09-232-479-13

Alignment Scores:
Pred. No.: 1.73 Length: 1362
Score: 78.00 Matches: 39
Percent Similarity: 45.24% Conservative: 37
Best Local Similarity: 23.21% Mismatches: 64
Query Match: 7.30% Indels: 28
DB: 4 Gaps: 10

US-10-024-955-7 (1-213) x US-09-232-479-13 (1-1362)
QY 40 lIeGlInSeRcIuThrIlleAspPrometLysValProAspHisAlaAspLysPheGlu 59
::: ||| |||::: |||
Db 304 GTGAGTCGAGAGTCACATGAGTCAGACACACCTCCGTCGGCTTATGACACTAAC 363
QY 60 ArgHisValGIYlLeValAspPheLysGIYlLeuAla--MetArgAsnIleGluAla 78
::: ||| |||::: |||
Db 364 GGCACCATTAAGCGCGCTGACCTTCCAAAGAGAGCTGAGTAAGTGAAGTGTACCTAC 423
QY 79 ArgGIYLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGIYlGluGlyIle 98
|||||
Db 424 AATGGGTG-----ATGCTGCACAGCCACATCAACGACAAATTGGG--- 468
QY 99 ValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMet-----GluTyr 116
::: |||::: |||
Db 469 -----AATGCTCTGTTAGGGGAAGGGGTACATGCTCTCCTACCGCTTACCCACATCATAT 519
QY 117 AspLeuAlaTyr--LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIle 135
||||| ||| ::||| |||
Db 520 GATCTTGGGTATGTGAGGCTGTGGTACCCCATTCGCCGTATAGGGCTT----- 567
QY 136 GlnAspPheValAlaLalaLeuSerLeuGlnIleSerAspGluLysAsnIle---ThrMet 154
||| ::| ::|
Db 568 ---GACCCAAATAATGTTAGCTACATCGACAGACAGTGAAGGCCGAGAGTACACCATATA 624
QY 155 ThrSerPheGluValArgGlnPheAlaAsnValAlaHisIleIleGlyLysLeuSerIle 174
||||| ::| |||||::: |||
Db 625 ACTGCAGCGGATGATTAACCAATTCATCATCAGACAGTACCACCA---GGTGGGGTAACAAATC 661
QY 175 LeuAspProIlePheGluValLeuSerAspValLeuThrAla----- 188
::: |||
Db 682 -----ACACTGTCTTCACAGCAACATTTGATGCTATCACAAGCCTCAGCATTTGGGGAGAG 735
QY 189 ---lIePheGlnAspThrValArg 195
::: |||
Db 736 CTCGTGTTTCAACAAGCGTCCAA 759
```

```
RESULT 26
US-09-784-990-13
; Sequence 13, Application US/09784990
; Patent No. 6464984
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, JEAN-CHRISTOPHE
; APPLICANT: BOUCHARON, ANNABELLE
; APPLICANT: RIVIERE, MICHEL
; TITLE OF INVENTION: AVIAN POLYNUCLEOTIDE VACCINE FORMULA
; FILE REFERENCE: 454313-2260
; CURRENT APPLICATION NUMBER: US/09/784,990
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 96/09339
; PRIOR FILING DATE: 1996-07-19
; EARLIER APPLICATION NUMBER: PCT/FR97/01326
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Infectious bursal disease virus
US-09-784-990-13

Alignment Scores:
Pred. No.: 1.73 Length: 1362
Score: 78.00 Matches: 39
Percent Similarity: 45.24% Conservative: 37
Best Local Similarity: 23.21% Mismatches: 64
Query Match: 7.30% Indels: 28
DB: 4 Gaps: 10

US-10-024-955-7 (1-213) x US-09-784-990-13 (1-1362)
QY 40 lIeGlInSeRcIuThrIlleAspPrometLysValProAspHisAlaAspLysPheGlu 59
::: ||| |||::: |||
Db 304 GTGAGTCGAGAGTCACATGAGTCAGACACACCTCCGTCGGCTTATGACACTAAC 363
QY 60 ArgHisValGIYlLeValAspPheLysGIYlLeuAla--MetArgAsnIleGluAla 78
::: ||| |||::: |||
Db 364 GGCACCATTAAGCGCGCTGACCTTCCAAAGAGAGCTGAGTAAGTGAAGTGTACCTAC 423
QY 79 ArgGIYLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGIYlGluGlyIle 98
|||||
Db 424 AATGGGTG-----ATGCTGCACAGCCACATCAACGACAAATTGGG--- 468
QY 99 ValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMet-----GluTyr 116
::: |||::: |||
Db 469 -----AATGCTCTGTTAGGGGAAGGGGTACATGCTCTCCTACCGCTTACCCACATCATAT 519
QY 117 AspLeuAlaTyr--LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIle 135
||||| ||| ::||| |||
Db 520 GATCTTGGGTATGTGAGGCTGTGGTACCCCATTCGCCGTATAGGGCTT----- 567
QY 136 GlnAspPheValAlaLalaLeuSerLeuGlnIleSerAspGluLysAsnIle---ThrMet 154
||| ::| ::|
Db 568 ---GACCCAAATAATGTTAGCTACATCGACAGACAGTGAAGGCCGAGAGTACACCATATA 624
QY 155 ThrSerPheGluValArgGlnPheAlaAsnValAlaHisIleIleGlyLysLeuSerIle 174
||||| ::| |||||::: |||
Db 625 ACTGCAGCGGATGATTAACCAATTCATCATCAGACAGTACCACCA---GGTGGGGTAACAAATC 681
QY 175 LeuAspProIlePheGluValLeuSerAspValLeuThrAla----- 188
::: |||
Db 682 -----ACACTGTCTTCACAGCAACATTTGATGCTATCACAAGCCTCAGCATTTGGGGAGAG 735
QY 189 ---lIePheGlnAspThrValArg 195
::: |||
Db 736 CTCGTGTTTCAACAAGCGTCCAA 759

RESULT 27
US-09-103-840A-2
; Sequence 2, Application US/09103840A
```

```

; CURRENT APPLICATION NUMBER: US/09/090,793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 13
; LENGTH: 19227
; TYPE: DNA
; ORGANISM: Vibrio marinus
US-09-090-793-13

```

```

Alignment Scores:
Pred. No.: 67.6      Length: 19227
Score: 78.50      Matches: 57
Percent Similarity: 37.34%      Conservative: 30
Best Local Similarity: 24.46%      Mismatches: 85
Query Match: 7.35%      Indels: 61
DB: 3      Gaps: 12

```

US-10-024-955-7 (1-213) x US-09-090-793-13 (1-19227)

```

QY 4 PheLeuLeuIleAlaAlaAla-----PheValAlaValSerAlaAspPro 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11814 TTTGCACTGGATCGCGCTTGTTCATCTTGTATAGCGGTAAAGTACGGTGTATAC 11873

QY 20 IleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAla 39
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11874 CTGCAT-----ACGGGTAAAGCCACACATGATGCTTGTGCGGATATCGCA 11921

QY 40 IleGluInserGluThrIleAspPrometLysVal----- 51
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11922 -----GCAGATCCTATGTTGCTGAATATAGGCTTTCGATATTCCAA 11963

QY 52 -----ProAspHisAlaAspLysPheGluArgHisValGlyIleValAsp 66
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11964 GCTTACCCAGCTAACAAATGATACATGCCCGTTTGACCAAAATTCACAAAGTCTA 12017

QY 67 PheLysGlyIleuLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12018 TTTGCCGGTGAAGCGCG-----GCCATGATGCTATTTGAAGCGT 12056

QY 87 GlnGlyAspAlaAsnValLysGlyGluGlu-----GlyIleValLysAlaHisLeuLeu 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12057 CAAAGTGATGACGATAGTGTGATCATATTACGCCATTATTAAAGCGCGCATTTA 12116

QY 105 Ile-----GlyValHisAspAspIleValSer-----MetGlu 115
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12117 TCGAATGACGGTAAAGCGAGTTTGTATTAAAGCCCGAACACCAAGGCCAAGTATTAGTA 12176

QY 116 TyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIleSerAspIle 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12177 TATGAACGCTTATGCCATGCAAGTGTGACCCGAGTACA----- 12218

QY 136 GlnAspPheValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThr 155
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12219 GTTAGCTATATT-----GAATGTCATGCAACGGCGCACACCTTAAGGTGAC 12263

QY 156 SerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeu 175
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12264 AATGTTGAATTCGTTGCGATGGAACCTTTTTCAGTCGGGTAAATTAACAAACCATTAAG 12323

QY 176 AspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArg 195
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12324 GGCTCGTTAAATCTAACCCTTGTGTCATTGTTAACCTGCCGCGTGTATGCTGACATGACC 12383

QY 196 LysGluMetThr-----LysValLeuAlaProAla 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12384 AAAGCTATGTAGCGCTAGCTAAGTAAAGCTTATTCCTGCA 12422

RESULT 24
US-09-090-793-12
; Sequence 12, Application US/09090793

```

```

; Patent No. 6140486
; GENERAL INFORMATION:
; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; FILE OF INVENTION: of polypeptide-like synthesis genes in plants
; FILE REFERENCE: CGME.131.0105
; CURRENT APPLICATION NUMBER: US/09/090,793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 12
; LENGTH: 40138
; TYPE: DNA
; ORGANISM: Vibrio marinus
US-09-090-793-12

```

```

Alignment Scores:
Pred. No.: 195      Length: 40138
Score: 78.50      Matches: 57
Percent Similarity: 37.34%      Conservative: 30
Best Local Similarity: 24.46%      Mismatches: 85
Query Match: 7.35%      Indels: 61
DB: 3      Gaps: 12

```

US-10-024-955-7 (1-213) x US-09-090-793-12 (1-40138)

```

QY 4 PheLeuLeuIleAlaAlaAla-----PheValAlaValSerAlaAspPro 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28797 TTTGCACTGGATCGCGCTTGTTCATCTTGTATAGCGGTAAAGTACGGTGTATAC 28856

QY 20 IleHisTyrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAla 39
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 28857 CTGCAT-----ACGGGTAAAGCCACACATGATGCTTGTGCGGATATCGCA 28904

QY 40 IleGluInserGluThrIleAspPrometLysVal----- 51
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28905 -----GCAGATCCTATGTTGCTGAATATAGGCTTTCGATATTCCAA 28946

QY 52 -----ProAspHisAlaAspLysPheGluArgHisValGlyIleValAsp 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28947 GCTTACCCAGCTAACAAATGATACATGCCCGTTTGACCAAAATTCACAAAGTCTA 29000

QY 67 PheLysGlyIleuLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29001 TTTGCCGGTGAAGCGCG-----GCCATGATGCTATTTGAAGCGT 29039

QY 87 GlnGlyAspAlaAsnValLysGlyGluGlu-----GlyIleValLysAlaHisLeuLeu 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29040 CAAAGTGATGACGATAGTGTGATCATATTACGCCATTATTAAAGCGCGCATTTA 29099

QY 105 Ile-----GlyValHisAspAspIleValSer-----MetGlu 115
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29100 TCGAATGACGGTAAAGCGAGTTTGTATTAAAGCCCGAACACCAAGGCCAAGTATTAGTA 29159

QY 116 TyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIleSerAspIle 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29160 TATGAACGCTTATGCCATGCAAGTGTGACCCGAGTACA----- 29201

QY 136 GlnAspPheValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThr 155
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29202 GTTAGCTATATT-----GAATGTCATGCAACGGCGCACACCTTAAGGTGAC 29246

QY 156 SerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeu 175
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29247 AATGTTGAATTCGTTGCGATGGAACCTTTTTCAGTCGGGTAAATTAACAAACCATTAAG 29306

QY 176 AspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArg 195
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29307 GGCTCGTTAAATCTAACCCTTGTGTCATTGTTAACCTGCCGCGTGTATGCTGACATGACC 29366

QY 196 LysGluMetThr-----LysValLeuAlaProAla 205

```



```

; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELE: <unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2019 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-621-855-3

Alignment Scores:
Pred. No.: 2.64 Length: 2019
Score: 78.50 Matches: 49
Percent Similarity: 39.63% Conservative: 37
Best Local Similarity: 22.58% Mismatches: 86
Query Match: 7.35% Indels: 45
DB: 4 Gaps: 11

US-10-024-955-7 (1-213) x US-09-621-855-3 (1-2019)
OY 18 AspproilehstyrAspIysIlethrgIu-----IleasnIysAla 32
DB 1072 GATACATTACAGACAGACTTATGGCAAAAAGATGCTTAATTCATCAGTAAACGG 1131
OY 33 IleasPaspAlaIleAlaIleIleGluInserGluThIleasPprometIysValPro 52
DB 1132 GTTAGACGTCGCCGTCAGAGGTTAA-----GATCC----- 1164
OY 53 AsphIsAlaAspIysPheGluArgHisValGlyIleValAspPheIysGlyIleuAla 72
DB 1165 -----AAACGACCAATGTTAGCTTATCTTCCTTGACCAACTGGT 1206
OY 73 MetArgAsnIleGlu---AlaArgGlyLeu---LysGlnMetIysArgGlnGlyAspAla 90
DB 1207 GTTGTAAACGTAATTAAGCTAGAGCTTATGATCAATGTTGGCATGATGATGGC 1266
OY 91 AsnValIysGlyGlu---GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
DB 1267 ATGATCCGTGTGACATGATGATTAATTAAGCAAAACACGACGACCGCATTTAGTGGT 1326
OY 107 Val-----HisAspPleValIserMetGluTyrAspLeuAlaTyr 120
DB 1327 GCTCCCTCCAGATATGTGTGTCATGATGTCGACATTAAGTAAAGTTAGACGT 1386
OY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAsp-----IleGln 136
DB 1387 AAA-----CCATATCTCTTAATTTTATTGATCAAAATGAAAAAGCTCAT 1431
OY 137 AspPheValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIlethrmethrSer 156
DB 1432 CCAGATGTAATTAATTAATTTATTAAGATGATGACATTTGACAGATTAACAAA 1491
OY 157 PheGluValArgGlnPheAlaAsnValVal-----AsnHisIleGlyGlyLeu 172
DB 1492 GGACGTACAGTGTGATTTCAAGAAATCAATTAATCATAGATCAACAGCTTGGGACAA 1551
OY 173 SerIleLeuAspProIlePheGlyValLeuSerAspValLeuThralaIlePheGlnAsp 192
DB 1552 GATTACAAAGATCAACGATTTGCTGATTCGGTTCAGAGTATGACAGATTAATGAA 1611
OY 193 ThrValArgIysGluMetThrIysValLeuAlaProAlaPheIysArgIu 209
DB 1612 ACAATTCGAAAAACGATGTTAAAGAAATTAATAATTCGTCACGAA 1662
RESULT 21

```

```

US-09-040-843-1
; Sequence 1, Application US/09040843
; Patent No. 6124119
; GENERAL INFORMATION:
; APPLICANT: Jaworski, Deborah J.
; APPLICANT: Wang, Min
; APPLICANT: Shilling, Lisa K.
; APPLICANT: Burnham, Martin
; APPLICANT: Fosberry, Andrew
; APPLICANT: Hodgson, John E.
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: Mech
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,843
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/057,535
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10082
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2599 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-040-843-1

Alignment Scores:
Pred. No.: 3.8 Length: 2599
Score: 78.50 Matches: 49
Percent Similarity: 39.63% Conservative: 37
Best Local Similarity: 22.58% Mismatches: 86
Query Match: 7.35% Indels: 45
DB: 3 Gaps: 11

US-10-024-955-7 (1-213) x US-09-040-843-1 (1-2599)
OY 18 AspproilehstyrAspIysIlethrgIu-----IleasnIysAla 32
DB 1510 GATACATTACAGAGACTTATGGCAAAAAGATGCTTAATTCATCAGTAAACGG 1569
OY 33 IleasPaspAlaIleAlaIleIleGluInserGluThIleasPprometIysValPro 52
DB 1570 GTTAGACGTCGCCGTCAGAGGTTAA-----GATCC----- 1602
OY 53 AsphIsAlaAspIysPheGluArgHisValGlyIleValAspPheIysGlyIleuAla 72
DB 1603 -----AAACGACCAATGTTAGCTTATCTTCCTTGACCAACTGGT 1644
OY 73 MetArgAsnIleGlu---AlaArgGlyLeu---LysGlnMetIysArgGlnGlyAspAla 90

```

DB 682 -----ACAGTCTCTCAGCTAACATTCGATGCGCATCACAAGCCTCAGCATCGGGAGAA 735
QY 189 ---IlePheGlnAspThrValArg 195
DB 736 CTCGTGTTCCAAACAAGCGTCCAA 759
RESULT 19
US-09-040-843-3
Sequence 3, Application US/09040843
Patent No. 6124119
GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah J.
APPLICANT: Wang, Min
APPLICANT: Shilling, Lisa K.
APPLICANT: Burnham, Martin
APPLICANT: Fosberry, Andrew
APPLICANT: Hodgson, John E.
APPLICANT: Lawlor, Elizabeth
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: Mecb
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040.843
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/057,535
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GML0082
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2019 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-040-843-3
Alignment Scores:
Pred. No.: 2.64 Length: 2019
Score: 78.50 Matches: 49
Percent Similarity: 39.63% Conservative: 37
Best Local Similarity: 22.58% Mismatches: 86
Query Match: 7.35% Indels: 45
DB: 3 Gaps: 11
US-10-024-955-7 (1-213) x US-09-040-843-3 (1-2019)
QY 18 AspproIleHisTyrAspLysIleThrGlu-----IleAsnLysAla 32
DB 1072 GATACATTACATGAGAGATTTATTTGGCAAAAAGATGCTTTAATTCAATCAGTAAGCG 1131
QY 33 IleAspAlaIleAlaIleGluGlnInserrGluThrIleAspPrometLysValPro 52

DB 1132 GTTAGACGTGCCCCGTGCAGGGGTTAAAA-----GATCCT----- 1164
QY 53 AsphIAlaAspLysPheGlnArgHisValGlyIleValAspPheLysGlyIleuAla 72
DB 1165 -----AAACGACCAATGTTGCTAGCTTATCTCTTGGCAACACTGCT 1206
QY 73 MetArgAsnIleGlu---AlaArgGlyLeu---LysGlnMetLysArgGlnLysAla 90
DB 1207 GTTGGTAAACAGTAATTTAGCTAGACCTTAGCTGAATCAATCTTTGGCGATGATATCG 1266
QY 91 AsnValLysGlyGlu---GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
DB 1267 ATGATCCGCTAGACATGAGTGAATTTATGAAAAACAGCAGTGAGCCGATTTAGTGGT 1326
QY 107 Val-----HisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
DB 1327 GCTCCTCCAGATATGTTGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1386
QY 121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAsp-----IleGln 136
DB 1387 AAA-----CCATATCTGTGAATTTTATTTGATGAATTTGAAAAACTCAT 1431
QY 137 AspPheValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSer 156
DB 1432 CCAGATGATTTAATATCTTATTTACAGTTTACATGATGATGATGATGATGATGATGATGAT 1491
QY 157 PheGluValArgGlnPheAlaAsnValAla-----AsnHisIleGlyGlyLeu 172
DB 1492 GGACGTACAGTTGATTTGAGATTCAGATTCATATATGATATGATGATGATGATGATGATGAT 1551
QY 173 SerIleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp 192
DB 1552 GAATTACAAAGATCAAGATTTCTGCTGATTCGTTCAAGATGATGATGATGATGATGATGAT 1611
QY 193 ThrValArgLysGluMetThrLysValLeuAlaProAlaPheLysArgGlu 209
DB 1612 ACAATTCGAAAAACGATGTATAAAGATTTAAAAAATTCATTCCTCCGTCAGAA 1662
RESULT 20
US-09-621-855-3
Sequence 3, Application US/09621855
Patent No. 6346608
GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah J.
APPLICANT: Wang, Min
APPLICANT: Shilling, Lisa K.
APPLICANT: Burnham, Martin
APPLICANT: Fosberry, Andrew
APPLICANT: Hodgson, John E.
APPLICANT: Lawlor, Elizabeth
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: Mecb
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/621,855
FILING DATE: 24-Jul-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/040.843

UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

Alignment Scores:
Pred. No.: 3.69 Length: 3123
Score: 79.50 Matches: 48
Percent Similarity: 38.31% Conservative: 47
Best Local Similarity: 19.35% Mismatches: 82
Query Match: 7.44% Indels: 71
DB: 1 Gaps: 10

US-10-024-955-7 (1-213) x US-08-094-889-2 (1-3123)

QY 18 ASPPTOLIEHS-----TYASPLYSILERTHLU-----GIU 28
DB 1205 GACCCACATTCCTTATGCAAGATGACCAAGAAAACAGACCTTCGAAGACAG 1264
QY 29 ILEASNLVLAILEASPSALALEA----- 38
DB 1265 CTCGCGAAGCAGTATGATGATATATCAGATTCCTTCGAGACCAATGTCATTCG 1324
QY 39 -----AlailegluInsergluThrIleasPromeTylsValProasPhisAlaasp 56
DB 1325 CTGCTTCATGAGAGCTCCCAAAAGCGGTATGAGAAAGAACTGAGAGAGTACGGCCAG 1384
QY 57 LYSPEGLIARGHISVALGLY---ILEVALASPHLEYSGLYLULEVALMETARGASN 75
DB 1385 GCTTCCTCGAGCATGCCAACAAGCTGTTGAGGTGCAAAATCTGCGCTTCATCTCA 1444
QY 76 ILEGLUALARGLYLEULYSGINMETLVS----- 85
DB 1445 AATATGAGAAGGTGTGAGAGTACGTATGCGACACAACAGATGATGCTGTGC 1504
QY 86 -----ArgGln 87
DB 1505 CCCGAGTAAATAATCTCGCTGACGTTGGCTGCCAGACCTCAGACCAAGTAGACACAG 1564
QY 88 GLYASPALAASNVALLYSGLYGLUGLYILEVALLYSALAHISLEULEUILLEGLYVAL 107
DB 1565 GACAACATGAGATGCTTTAAGATCAGTGGAGAAACAAGTGGAGTCTCACTGAAGCA 1624
QY 108 HISASPSALLEVALSERMETGLUYTRASPLEUALATYLLYSLEUGLYASPLEUHIISPRO 127
DB 1625 GTTGTGACATCACTTCAGTGTGATGATTCCTCTGTTTCA----- 1666
QY 128 THRTHHISVALILESERASPILEGINAASPHLEVALALEUSERLEUGLIULESER 147
DB 1667 GAAATACATATCTGGAATGGAACAATGCTGATTCCT-----CTCCAAAGGCA 1720
QY 148 ASPLGLYASNILETHRMETHRSERPHEGLIUALARGGLNPHLEALASNVALYLAASN 167
DB 1721 GATGTGCAATACCTGATGAGAACTGCTGGGCCATCCGAGGCCGTGACACCAAGATCATT 1780
QY 168 HISILEGLYGLYLEUSERILELEUASPROILIEPHEGLYVALLEUSERSPVALLEUTHR 187
DB 1781 CACATC-----ATTAAATGCAAGATGAGAAACATATGAAGACT 1816

QY 188 ALAILEPHEGLINSPHRVALARGLYSGIMETTHLYSVALLEUALA----- 203
DB 1817 GGAGTTTATACCGAGAGCTA---CTGGAGACCAACCAACTGCTCTCTGAACAGTTANG 1873
QY 204 PROALAPHELYSARGGLULEUGLI 211
DB 1874 CCACGTTTGTGTAACAAGTTGAG 1897

RESULT 18
US-09-450-433-1
Sequence 1, Application US/09450433
Patent No. 6468984
GENERAL INFORMATION:
APPLICANT: Innovo Biotechnologies Ltd.
APPLICANT: Aboud-Pirak, Esther
APPLICANT: Monadeev, Limor
APPLICANT: Pirak, Michael
APPLICANT: Shaoul, Esther
TITLE OF INVENTION: DNA vaccine for protecting an avian against infectious bursa
FILE REFERENCE: P-2181-US
CURRENT APPLICATION NUMBER: US/09/450,433
CURRENT FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: US 60/138,093
PRIOR FILING DATE: 1999-06-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 1365
TYPE: DNA
ORGANISM: Infectious bursal disease virus
US-09-450-433-1

Alignment Scores:
Pred. No.: 1.3 Length: 1365
Score: 79.00 Matches: 39
Percent Similarity: 45.24% Conservative: 37
Best Local Similarity: 23.21% Mismatches: 64
Query Match: 7.40% Indels: 28
DB: 4 Gaps: 10

US-10-024-955-7 (1-213) x US-09-450-433-1 (1-1365)

QY 40 ILEGLUINSERGLUTHRILEASPROMETLYSVALPROASPHISALASPLYSPEGLU 59
DB 304 GTGAGTCGAGAGTCTCAGATGAGAGTCAACACACACTCCCTGGTGGCTTATGACACTAAT 363
QY 60 ARGHISVALGLYILEVALASPHLEYSGLYLULEUALA---METARGASNILEGLUALA 78
DB 364 GGCACCATTAAGCGCGCTGACCTCCAAAGAGCTGAGTGAAGTCAAGATGTTAGCTAC 423
QY 79 ARGGLYLEULYSGINMETLYSARGGLINLYASPSALASNVALLYSGLYLULEGLYILE 98
DB 424 AATGGGTTG-----ATGTCGCAACAGCCCAACATCAACACAAATCGGG--- 468
QY 99 VALLYSALAHISLEULEUILLEGLYVALHISASPSALLEVALSERMET-----GLUTYR 116
DB 469 -----AACGTCCTATAGGGGAAGGGGTAAACCGTCTCAAGCTTACCACATCATAT 519
QY 117 ASPLEUALATYR---LYSLEUGLYASPLEUHIISPROTHRTHHISVALILESERASPILE 135
DB 520 GATCTGGGTATGTGAGACTCGGTGACCCCATTCGCCGTATAGAGGCTC----- 567
QY 136 GLNASPHLEVALALEALEUSERLEUGLIULESERASPLGLYASNILE---THRMET 154
DB 568 ---GACCCAAAATGTTACCAACATGTGACACAGTGAAGAGGCCAGAGTCTACACCATTA 624
QY 155 THRSPHEGLIUALARGGLNPHLEALASNVALYLAASNHISILEGLYGLYLEUSERILE 174
DB 625 ACTGCAGCGGATGATTAACCAATTTCTATCACAAGTAC---CAAGAGGTGGGGTAACAATC 681
QY 175 LEUASPROILIEPHEGLYVALLEUSERASPSVALLEUTHRALA----- 188

Qy 198 MetHrLysValLeuAlaProAlaPheLysArgGluLeuLys 212
|||:|||||
Db 889 GCGGTTACCGTATCTCTGAAGAGATCGGTATGAGCTGGAAAA 933

RESULT 16
US-09-472-971-7
; Sequence 7, Application US/09472971
; Patent No. 6197547
; GENERAL INFORMATION:
; APPLICANT: SOGO, Kazuyo
; APPLICANT: YANAGI, Hideki
; APPLICANT: YURA, Takashi
; TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
; FILE REFERENCE: 1422-409P
; CURRENT APPLICATION NUMBER: US/09/472,971
; CURRENT FILING DATE: 1999-12-28
; EARLIER APPLICATION NUMBER: JP10-372965
; EARLIER FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 4524
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-472-971-7

Alignment Scores:
Pred. No.: 3.51 Length: 4524
Score: 81.50 Matches: 48
Percent Similarity: 36.17% Conservative: 37
Best Local Similarity: 20.43% Mismatches: 85
Query Match: 7.63% Indels: 65
DB: Gaps: 10

US-10-024-955-7 (1-213) x US-09-472-971-7 (1-4524)

Qy 13 VALAlaValSerAlaAspProIleHisTyrAspLysIleThrGluLutLeuAsnLysAla 32
|||:|||||
Db 701 GTTGCTCGCGGCATGACCCGATG-----GACCTGAAGCTGTGT 739

Qy 33 IleAspAspAlaIle--AlaAlaIleGluInserGluThrIleAspProMetLysVal 51
|||:|||||
Db 740 ATCGACAAGCGGTTACCGCTCGACGTGAAGA-----CTGAAGCCGTGCCGTA 790

Qy 52 ProAspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyLutLeu 71
|||:|||||
Db 791 CCATCTCTGACTCTAAAGCGATTGCTCAGTTGTTACCATCTCCGTAACCTCCGACGAA 850

Qy 72 AlaMetArgAsnIleGluAlaArgLysLeuLysGlnMetLysArgGluLysAlaAsn 91
|||:|||||
Db 851 ACCGTAGTAAGCTGATCGCTACACGATGACAAAGTC----- 889

Qy 92 ValLysGlyGluGluGlyIleValLysAlaHisLeuLeuIleGlyValHisAspAsp--- 110
|||:|||||
Db 890 -----GGTAAGAAGCGGTTATCAACCGTTGAAGCGGTAACCGGCTCGACGACGACTG 943

Qy 111 -----IleValSerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeu-----His 126
|||:|||||
Db 944 GACGTGGTTGAAGTATGATTCGACCGTGGCTACTCTCTCATCTTCAACAAG 1003

Qy 127 ProThrThrHisValIleSerAspIleGlnAspPheValValAlaLeuSerLeuGluIle 146
|||:|||||
Db 1004 CCGGAACCTGGCGAGTAGAACGTGAAGACCCGTTCAATCTGCTGCGTGAACAAGAAATC 1063

Qy 147 SerAsp----- 148
|||:|||||
Db 1064 TCCAACATCCGGAATAGTGCCTCCGCTTCGAAGCTGTGCCAAGCAGCAACCGCTG 1123

Qy 149 -----GluGlyAsnIleThrMetThrSerPheGluValArgGln 161
|||:|||||
Db 1124 CTGATCATCGCTGAAGATGTAGAAGCGCAAGCGCTGCAAACTGCT----- 1168

Qy 162 PheAlaAsnValValAsnHisIleGlyGlyLeu-----SerIleLeuAspPro 177
|||:|||||
Db 1169 -----GTTGTTAAACACCATTCGTGGCATTCGGAAGAGTCGCGGTTAAAGCACCG 1219

Qy 178 IlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGlu 197
|||:|||||
Db 1220 GCCTTCGGC-----GATCGTCTGAAGCTATGCTGACGAGATTCGAACCTGACT 1270

Qy 198 MetHrLysValLeuAlaProAlaPheLysArgGluLeuLys 212
|||:|||||
Db 1271 GCGGTTACCGTATCTCTGAAGAGATCGGTATGAGCTGGAAAA 1315

RESULT 17
US-08-094-889-2
; Sequence 2, Application US/08094889
; Patent No. 5470966
; GENERAL INFORMATION:
; APPLICANT: Shinji HIRANO et al.
; TITLE OF INVENTION: NEURAL -CATENIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/094,889
FILING DATE: July 22, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3123 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:


```

DB 1948 GGTGCTACTGTGACCTTCGTAATGATATTATGATCTTCAATGAGAGCTCA 2007
OY 173 SerIleuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp 192
DB 2008 GAATTACAGACCAACACCTTGTCTGCTTTGGAGGCTTCACAGAGTACTGACTACGAA 2067
OY 193 ThrValArgLysGluMetThrLysValLeuAlaProAlaPheLysArgLys 209
DB 2068 ACTGTCAGAAACCAATGATGAAAGAAATTAAATAATTCATTCGACACAGAA 2118

RESULT 14
US-09-134-001C-1220
; Sequence 1220, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1220
; LENGTH: 1089
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1220

Alignment Scores:
Pred. No.: 0 391 Length: 1089
Score: 82.00 Matches: 31
Percent Similarity: 45.67% Conservative: 27
Best Local Similarity: 24.41% Mismatches: 39
Query Match: 7.68% Indels: 30
DB: 4 Gaps: 7

US-10-024-955-7 (1-213) x US-09-134-001C-1220 (1-1089)
OY 84 MetLysArgGlnGluLysPalaAsnValLysGlyGluGluGlyLeuValLysAlaHisLeu 103
DB 646 GTGAACGTGAA-----AAAGGTCAAGAGCATTA-----675
OY 104 LeuIleGlyValHisAspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGly 123
DB 676 -----GATATTTGAATCAACTTATGATTTAGATTCCTACACGC 717
OY 124 AspLeuHisProThr-----ThrHisValIleSerAspIleGlnAspPheValAlaLeu 142
DB 718 GTTATCATTCACCACTCAATCCCAT-----AGTGAATATGATACATTAATTAATTAATTA 771
OY 143 SerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPhe 162
DB 772 GCACACACAGAT-----CATGCACATGTGATTAACGACGATTTTAATTTA-----816
OY 163 AlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGlyValLeu 182
DB 817 ---AAATAACTATGTCACGTTCAAGAAATACAGACACTCAAC-----GTT 858
OY 183 SerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLysValLeu 202
DB 859 AATGATTTTATCGACAGCAACCACTAATGTATCATCAAGGACCACTTAAGTATTTTA 918
OY 203 AlaProAlaPheLysArgLys 209
DB 919 TTAAGCAAGATAGCTAAAGAG 939

RESULT 15
US-09-472-971-5

```

```

; Sequence 5, Application US/09472971
; Patent No. 6197547
; GENERAL INFORMATION:
; APPLICANT: SOGO, Kazuo
; APPLICANT: YANAGI, Hideki
; APPLICANT: YURA, Takashi
; TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
; FILE REFERENCE: 1422-409P
; CURRENT APPLICATION NUMBER: US/09/472,971
; PRIOR FILING DATE: 1999-12-28
; EARLIER APPLICATION NUMBER: JP10-372965
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1647
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-472-971-5

Alignment Scores:
Pred. No.: 0 821 Length: 1647
Score: 81.50 Matches: 48
Percent Similarity: 36.17% Conservative: 37
Best Local Similarity: 20.43% Mismatches: 85
Query Match: 7.63% Indels: 65
DB: 4 Gaps: 10

US-10-024-955-7 (1-213) x US-09-472-971-5 (1-1647)
OY 13 ValAlaValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAla 32
DB 319 GTTGTGCGGCGCATGAACCGATG-----GACCTGAACGCTGT 357
OY 33 IleAspAspAlaIle-----AlaAlaIleGluGlnSerGluThrIleAspProMetLysVal 51
DB 358 ATCGACAAACCGGTTCACCGCTGCGAGTTGAAGAA-----CTGAAGCGCTGTCCGTA 408
OY 52 ProAspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyLeu 71
DB 409 CCATGCTGTGACTTAAGAGGATTCGACAGTGTGTAACCATCCGCTAATCCGACGAA 468
OY 72 AlaMetArgAsnIleGluAlaArgGlyLeuLysGluMetLysArgGlnGlyAspAlaAsn 91
DB 469 ACCGTAGGTAACTGATCCCTGAACGATGACAAAGTC-----507
OY 92 ValLysGlyGluGlnGlyIleValLysAlaHisLeuLeuIleGlyValHisAspAsp---110
DB 508 -----GTTAAAGAGCGCTTATCACCGCTGAAGACGTAACCGTGCAGAGCAACTG 561
OY 111 -----IleValSerMetGluTyrAspPheValTyrLysLeuGlyAspLeu-----His 126
DB 562 GACGTGTTCAAGTATGACGTTGCGACCGTGTACTCTGTCTTACTTCAATCAACAAAG 621
OY 127 ProThrThrHisValIleSerAspIleGlnAspPheValAlaLeuSerLeuGluIle 146
DB 622 CCGAAACGTGCGCAGTAGAAGTGAAGAACCCGTTCACTCTGCTGCTGACAAAGAAATC 681
OY 147 SerAsp-----148
DB 682 TTCACATCCCGCAAAATGCTGCCGGTTCTGGAAGCTGTGTCACAAACGCAACCGCTG 741
OY 149 -----GluGlyAsnIleThrMetThrSerPheGluValArgGln 161
DB 742 CTGATCATCGCTGAAGATGTAAGGAGGCAAGCGCTGGCAACTGCT-----786
OY 162 PheAlaAsnValValAsnHisIleGlyGlyLeu-----SerIleLeuAspPro 177
DB 787 -----GTTGTTAACACACATTCGTGCATCGTGAAGATCCGTCGCTTAAGACACCG 837
OY 178 IlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGlu 197
DB 838 GGCCTCGGC-----GATCGTGTGAAGCTATGCTGAGATTCGCAACCCCTGACT 888

```

```

; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: Ip90
; INDIVIDUAL ISOLATE: Isolate from I. persulcatus from
; INDIVIDUAL ISOLATE: Soviet Union
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 125..949
; OTHER INFORMATION: /product= "Ospsa"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 959..1843
; OTHER INFORMATION: /product= "Ospsb"
;
US-08-479-017-7

```

```

Alignment Scores:
Pred. No.: 0.439 Length: 1958
Score: 84.50 Matches: 52
Percent Similarity: 40.77% Conservative: 43
Best Local Similarity: 22.32% Mismatches: 87
Query Match: 7.91% Indels: 51
DB: Gaps: 12

```

US-10-024-955-7 (1-213) x US-08-479-017-7 (1-1958)

```

QY 1 MetMetLysPheLeu-----LleAlaAlaValAlaPheValAlaVal 15
DB 959 ATGAAAAAATATTACTAGGATTGCTTGTAGTATTAGCTTTAATACAGTGACAAAAA 1018
QY 16 SerAlaAspProIleHisTyrAspLysLleThgIuLleAsnLys----- 31
DB 1019 GGTGCTGAGCCAAACACATGATCAAGACGTTGAAGACTTAATAAAAGATCAAAAAAGC 1078
QY 32 -----AlaIleAspAspAlaLleAlaLleAlaLleGluInSer 43
DB 1079 GACTTAAAAAAGATCTCTCTTGTGTACAGACAGACGCTGAAGTTATTATTAACAAT 1138
QY 44 GluThrIleAspProMetLysValProAspHisAlaAspLysPheGluThrHisValGly 63
DB 1139 GAA---ATTTCATCAAGCAAAAGAAAAAAGAACGATAAATATGAAATTAAAGA---TCA 1192
QY 64 IleValAspPheLysGlyLleuAlaMetArgAsnIleGluAlaArgGlyLeuLysGln 83
DB 1193 ATAGTGTGAC-----AAGCTTACGCTTAAAGCCTTTCTGAG 1228
QY 84 MetLysArgGlnIlyAspAlaAsnValLysGlyGluGluIlyleValLysAlaHisLeu 103
DB 1229 ---AAGAATACTGCTGCTGAGAGCTTGAAGGTTTAAAGCT---GACAAAAACAAAGTA 1282
QY 104 LeuIleGlyValHisAspAspIle-----ValSerMetGlu---TyrAsp----- 117
DB 1283 ACAATGTGGTTTCTGACGATCTTAATACATATACATATGAACATATGATGCAAGCAAC 1342
QY 118 -----LeuAlaTyrLysLeuGlyAspLeuHisProThrHisVal 131
DB 1343 AAAAAAATTTCAAGCCAAAGGCTTAAAAAACAGGATCACTTAACAGAAACTTTACAAA 1402
QY 132 IleSerAspIleGlnAspPheValValAlaLeuSerLeuGluIleSerAspGluGlyAsn 151
DB 1403 ACTAGTAATAATTAGC-----GCAAAAGAAATTAACAAGATCAAAATAT 1444
QY 152 IleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIle----- 169
DB 1445 ACTACAAATGAATATACAGAAATGACAGACGCTGACAAATGCTTCAAAAGCAGTGAAACT 1504

```

```

QY 170 -----GlyIleuSerIleLeuAspProIlePheGlyValLeuSerAspValLeuThr 187
DB 1505 CTAATAAATGATATATCATCCTTGAAGAGACTCTTGTATGAGAAAAACACCTTAACATA 1564
QY 188 AlaIlePheGlnAspThrValAlaArgLysGluMetThrLys 200
DB 1565 AAAAGAGGCACGTCTTCTTAAAAAAGAAATGAAAAA 1603
RESULT 13
US-09-134-001C-1244
; Sequence 1244, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1244
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1244

```

```

Alignment Scores:
Pred. No.: 0.613 Length: 2472
Score: 84.50 Matches: 50
Percent Similarity: 39.17% Conservative: 35
Best Local Similarity: 23.04% Mismatches: 87
Query Match: 7.91% Indels: 45
DB: Gaps: 11

```

US-10-024-955-7 (1-213) x US-09-134-001C-1244 (1-2472)

```

QY 18 AspProIleHis-----TyrAspLysLleThgIuLleAsnLysAla 32
DB 1528 GATACACTTCATTAACGCTGTCATTTGACAAACAGATGCTGTCAATTCATGTAGTAAGCT 1587
QY 33 IleAspAspAlaLleAlaLleGluInSerGluThrIleAspProMetLysValPro 52
DB 1588 GTTAGAAGAGCTGCTGCTGCTTAA-----GATCA----- 1620
QY 53 AspHisAlaAspLysPheGluThrHisValGlyIleValAspPheLysGlyLleuAla 72
DB 1621 -----AAAGCTCCAAACGGTAGTTTATTTCTTAGAGACTACAGGT 1662
QY 73 MetArgAsnIleGlu---AlaArgGlyLeu---LysGlnMetLysArgGlnGlyAspAla 90
DB 1663 GTGCGTAAACTGAATATGCTGCTGCTTAAAGCTGATCTATGTTGCGTGAAGACGATGCA 1722
QY 91 AsnValLysGlyLyu---GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
DB 1723 ATGATTCGCTAGATATGACTGAATTTATGGAACAAACATGCTCAGTCGATTAGTGGT 1782
QY 107 Val-----HisAspAspIleValSerMetGluIuTyrAspLeuAlaTyr 120
DB 1783 GCACCTCCAGATATGAGACATGATGACGGCGTCATTTGACGAAAAAGTTAGACGT 1842
QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAsp-----IleGln 136
DB 1843 AAA-----CCATACCTCTGTCATTTTATTTGATGCAAAATTGAGAAAGCACAT 1887
QY 137 AspPheValValAlaLeuSerLeuGluIleSerAspGlnGlyAsnIleThrHisIle 156
DB 1888 CTTGACGTATTTATATCTTCTTACAAAGTTTATGATGATGCTATTTAAACAGTACTAAA 1947
QY 157 PheGluValArgGlnPheAlaAsnValVal-----AsnHisIleGlyGlyLeu 172

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/137,175A
FILING DATE: 26-OCT-1993
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08972
FILING DATE: 22-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: BARBOUR-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: Ip90
INDIVIDUAL ISOLATE: Isolate from I. persulcatus from
INDIVIDUAL ISOLATE: Soviet Union
FEATURE:
NAME/KEY: CDS
LOCATION: 125..949
OTHER INFORMATION: /product= "Ospa"
FEATURE:
NAME/KEY: CDS
LOCATION: 959..1843
OTHER INFORMATION: /product= "Ospb"
US-08-137-175A-7

Alignment Scores:
Pred. No.: 0.439 Length: 1958
Score: 84.50 Matches: 52
Percent Similarity: 40.778 Conservative: 43
Best Local Similarity: 22.328 Mismatches: 87
Query Match: 7.91% Indels: 51
DB: 1 Gaps: 12

US-10-024-955-7 (1-213) x US-08-137-175A-7 (1-1958)
QY 1 MetMetLysPheLeuLeu-----IleAlaAlaValAlaPheValAlaVal 15
DB 959 ATCAAAAAATATTACTAGCATTTGCTTAGTATTAGCTTAATAGCATGTGACAAAAA 1018
QY 16 SerAlaAspProIleHisTyrAspLysIleThcGluGluIleAsnLys----- 31
DB 1019 GGTGCTGACCAACAAACAAATGATCAAGCGTTGAAAGACTTAAAAAAGATCAAAAAAGAC 1078
QY 32 -----AlaIleAspAspAlaIleAlaIleAlaIleGluGlnSer 43
DB 1079 GACTCTAAAAAAGATCTCTTGTGTAAACAGACACGCGTGAAGTATTATTATAACAAAT 1138
QY 44 GluThrIleAspProMetLysValIProAspHisAlaAspLysPheGluThrHisValGly 63

```

```

DB 1139 GAA---ATTTTCATCAGCAAAAGAAAAAATGAAGCAGTAATATGAAATTAGA---TCA 1192
QY 64 IleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGln 83
DB 1193 ATAGTGGAC-----AAGTTGACCTTAAGGCCCTTTCTGAG 1228
QY 84 MetLysArgGlnGlyAspAlaAsnValLysGlyGluGluLysIleValLysAlaHisLeu 103
DB 1229 ---AAGAAATCTGCTGTCGACAGCTTGAAGCGTTTAAAGCT---GACAAAACAAAGTA 1282
QY 104 LeuIleGlyValHisAspSple-----ValSerMetGlu---TyrAsp----- 117
DB 1283 ACAATGTGGTTCTGACGCTTCAATTAACAAATATATAGAACATATGATCCAAAGCAAC 1342
QY 118 -----LeuAlaTyrLysLeuGlyAspLeuHisProThrHisVal 131
DB 1343 AAAAAAATTCAGCCCAAGCGCTTAAAAAAGCAGGATCACTAACAGAAACTTTCAAA 1402
QY 132 IleSerAspIleGlnAspPheValAlaIleSerLeuGluIleSerAspGluLysAsn 151
DB 1403 ACTAGTAATAATTAGC-----GCAAAAGAAATTAACAAGATCAATTAAT 1444
QY 152 IleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIle----- 169
DB 1445 ACTACAAATGAAATATACAGAAATGACAGACGCTGACAAATGCTTCAAAAGCAGTGAAGT 1504
QY 170 -----GlyGlyLeuSerIleLeuAspProIlePheGlyValLeuSerAspValLeuThr 187
DB 1505 CTAAAAAATGATATCAACCTTGACAGAGACTGTTGAGTGGAAGAAACAACTTAACAAATA 1564
QY 188 AlaIlePheGlnAspThrValArgLysGluMetThrLys 200
DB 1565 AAGAGCGCAGCTGTACTTTAAAAAAGAAATGAAAAA 1603

RESULT 12
US-08-479-017-7
Sequence 7, Application US/08479017
Patent No. 6143872
GENERAL INFORMATION:
APPLICANT: BARBOUR, Alan G.
APPLICANT: BERGSTROM, Sven
APPLICANT: HANSSON, Lennart
TITLE OF INVENTION: IMPROVEMENT IN BORRELLIA BURGDOERFERI AND
TITLE OF INVENTION: PROPHYLAXIS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,017
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/137,175
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: PCT/US92/08972
FILING DATE: 22-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: BARBOUR-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528

```


Pred. No.: 0.157 Length: 2157
 Score: 88.50 Matches: 30
 Percent Similarity: 45.08% Conservative: 25
 Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.29% Indels: 17
 DB: 1 Gaps: 3

US-10-024-955-7 (1-213) x US-08-336-618-25 (1-2157)

QY 18 AspproilehstYrAspYsIleThrGluGluIleAsnIleAspAlaIle 37
 DB 322 GACAAATTCCTTGGAGTGGAAAAGGAGGATCATCAGCTTGACATTCACATA 381
 QY 38 AlaAlaIleGluInserGluThrIle----- 46
 DB 382 GCCACCATGAAGCTGGGGAGGTGTGCCCATCACCCTGCACCAACAGAAATATGCTACGCT 441
 QY 47 -----AspprometlyValProAspHisAlaAsp--LysPheGluAlaArgHisVal 62
 DB 442 TCAGCAGGACAGCTCTCCAAAGATTCCTCCCAATGCCACCTGTATTTGAG-----GTG 495
 QY 63 GlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLys 82
 DB 496 GAGTTGTTGAGTTTAAAGGAGAGATCTGTACGAGAGAGAGAGATGGCGGAATCATTCGC 555
 QY 83 GlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLysAlaHis 102
 DB 556 AGAATACAGACCTCGGGGTAAAGCTATGCTAAGCCCAATGAGAGGTGCTATCGGAGCTT 615
 QY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122
 DB 616 GCACGTGAAGGACTACTACAAAGCAAGCTCTTGAACAGCGGAGCTCGCTTGGAGATT 675
 QY 123 GlyAsp 124
 DB 676 GCGGAG 681
 RESULT 8
 US-08-002-024B-7/c
 ; Sequence 7, Application US/08002024B
 ; Patent No. 5798103
 ; GENERAL INFORMATION:
 ; APPLICANT: MOOI, Frederik R
 ; TITLE OF INVENTION: WHOPPING COUGH VACCINE
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: YOUNG & THOMPSON
 ; STREET: 745 South 23rd Street
 ; CITY: Arlington
 ; STATE: VA
 ; COUNTRY: US
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/002,024B
 ; FILING DATE: 08-JAN-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP 92200038.5
 ; FILING DATE: 08-JAN-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: PATCH, Robert J
 ; REGISTRATION NUMBER: 17,355
 ; REFERENCE/DOCKET NUMBER: BO-37424
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703/521-2297
 ; TELEFAX: 703/585-0573
 ; TELEX: 248425 EMBON
 ; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1155 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: mat-peptide
 ; LOCATION: 112..1128
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1128
 ; NAME/KEY: sig-peptide
 ; LOCATION: 1..111
 ; US-08-002-024B-7
 Alignment Scores:
 Pred. No.: 0.0855 Length: 1155
 Score: 87.50 Matches: 33
 Percent Similarity: 36.67% Conservative: 33
 Best Local Similarity: 18.33% Mismatches: 61
 Query Match: 8.19% Indels: 53
 DB: 1 Gaps: 6

US-10-024-955-7 (1-213) x US-08-002-024B-7 (1-1155)

QY 14 AlaValSerAlaAspProilehstYrAspYsIleThrGluGluIleAsnLysAlaIle 33
 DB 654 GCGCTCTGGCTGGTCCCGCGCATGAA-----AACCATTTGTA 616
 QY 34 AspAspAlaIleAlaIleGluInserGluThrIleAspPrometlyValProAsp 53
 DB 615 GGTGAACAGGTTCCGAAACTGCGACGACACAGCTTGGCGCGCTGTGAGTGCGCCGAC 556
 QY 54 HisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyGluLeuAlaMet 73
 DB 555 ---CTTGATCGCCTGATAGCGAACCCTCCGCTTGTGCGCTTGATGTGACGACCTCGAA 499
 QY 74 ArgAsnIleGluAlaArgGlyLeuLysGlnMet-----LysArg 86
 DB 498 AAGCATTCGCTTCTGTGGGATACAGCACCTGCGCGGCTTCCGACCTTGTGATGGCGGA 439
 QY 87 GlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys----- 100
 DB 438 ACTGCCAAATCGATATAGATGCGATGCGGCTCTCCGAGGAGCGCAACGCAAGTC 379
 QY 101 -----AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAsp 117
 DB 378 GACACGATGCCCGGATGGCATAGGACTGCGCGCGGATCTTGTA----- 334
 QY 118 LeuAlaTyrLysLeuGlyLysPheHisProThrHisValIleSerAspIleGlnAsp 137
 DB 333 -----GTTACTGCCAGCTTCCACACCATTCGACCG 301
 QY 138 PheValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPhe 157
 DB 300 TTCGCCGTCATGCTGTCTCGTCGAGTC----- 274
 QY 158 GluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspPro 177
 DB 273 -----TTCCTGTATGTGTGCGGACCTTGACGAGTGTGAGACGAGGACCG 226
 RESULT 9
 US-09-103-840A-1/c
 ; Sequence 1, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

```

: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961,527
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 79:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5066 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: US-08-961-527-79
:
Alignment Scores:
Pred. No.: 0.463 Length: 5066
Score: 89.00 Matches: 52
Percent Similarity: 38.46% Conservative: 38
Best Local Similarity: 22.22% Mismatches: 100
Query Match: 8.33% Indels: 44
DB: 4 Gaps: 10
:
US-10-024-955-7 (1-213) x US-08-961-527-79 (1-5066)
Qy 7 Ileaalaavalaalaphavalalavalseralaaspproilehistryasplysilethr 26
Db 1804 ATTATGGCCCTAACCAATATTCATTCATGAGAGCTCCC--AATGTTGGAAATCAACC 1748
Qy 27 GIUGIUIleasnlyslaleleaspsalalealealelelelelelelelelelelele 46
Db 1747 --CTATTTAACTCGATGCTGTGAGCAATCTCCATTTGTAAGAAGATGTCGAAGAGTG 1691
Qy 47 Aspprometlyslvalprcasphlsalsaplysphegiuarghlsvalglylevalasp 66
Db 1690 ACACCTGACCCGATTTATGCAACAGCGGTGAGTGCCTCAATCTCTTTTAAGCATGATGAT 1631
Qy 67 Pheylsglygluleualamelargasnilleglualargglyleulysglinmetlysarq 86
Db 1630 ACAGAGAGA-----ATTGATGATGTCGATGCTCTTCATGATGACAAATCAAGCAC 1580
Qy 87 Glnglyaspalaasn-----Vallysglygluclu 96
Db 1579 CAGGCAGAAATTCATGATGAGAACACAGATGTTATCGTTTGTGCTGTCGTGAAGGAA 1520
Qy 97 Glyllevallys-----Alahlsleulellelelyvalhlspsaplle 111
Db 1519 GGAATTAAGTATGATGACAGCAATAGCTAGCTGTAAGCTTTATTAAGACCAACACAGTT 1460
Qy 112 Val-----Sermetglutyaspleu----- 118
Db 1459 ATCCCGCAGTCACAAAGGTGACAAACCTGAGATGAGAAATGATATATATGATTTCTAT 1400
Qy 119 AlaTyrlYslsleuglyAspleuhisProthrhHisVal-----Ileserasp 134
Db 1399 GCTCTCGGTTTGTTGGTGAACCATTTGCTATCTCATCTGTCATGTAATCGGTACAGGGAT 1340
```

```

Qy 135 Ileglnasphvalvalaleleuserleugliuleseraspjluglyasnillethmet 154
Db 1339 GTGCTAGATGGATGCTGAAGAAATCTTCCAAATGATATGAGGAAGAAATCCAGATGTC 1280
Qy 155 Threserphlegluvalarglmphealalsnval-----Valasnhlsile 169
Db 1279 ATTAACTTTACTGATGATGTCGTCTTAACGTTGGAAATCAAGCTTGATCAATGCTATC 1220
Qy 170 Glyglyleuser-----Ileleuaspproilephleglyvalleuseraspvalleuthr 187
Db 1219 TTGGGAGAAGACCGTGTATTGCTAGTCTGCTGGAACAACCTCGTATGCCATTTGAT 1160
Qy 188 Alailepheginaspphrvalarglysglumethrlyslval 201
Db 1159 ACCCACTTACAGATACAGATGCTCAAGAGTTTACCATGATT 1118
:
RESULT 7
US-08-336-618-25
: Sequence 25, Application US/08336618
: Patent No. 5763590
:
: GENERAL INFORMATION:
: APPLICANT: Peattie, Debra A.
: APPLICANT: Harding, Matthew W.
: APPLICANT: Livingston, David J.
: TITLE OF INVENTION: ISOLATION OF AN Mr 52,000 FK506 BINDING
: TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
: TITLE OF INVENTION: cDNA
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P. C.
: STREET: Two Millitia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/336,618
: FILING DATE: 09-NOV-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/963,325
: FILING DATE: 16-OCT-1992
: APPLICATION NUMBER: US 07/777,752
: FILING DATE: 11-OCT-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/
: FILING DATE: 09-OCT-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: VPI91-06A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2157 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 100..1476
:
: US-08-336-618-25
:
Alignment Scores:
```

FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(5361)
US-08-973-462-2

Alignment Scores:
Pred. No.: 0.375 Length: 5361
Score: 90.00 Matches: 47
Percent Similarity: 44.55% Conservative: 43
Best Local Similarity: 23.27% Mismatches: 78
Query Match: 8.43% Indels: 34
DB: Gaps: 9

US-10-024-955-7 (1-213) x US-08-973-462-2 (1-5361)

QY 22 TYRASPPLYSILETHRGUUGLUILEASNLYSALAILLEASPSAPALAILLEALAILLEGLU 41
DB 2935 TTTTAATCTGTATTAGATTAAGTAGAGAAACGTGAAATTTAGCGAGAAATTTAGAA 2994
QY 42 GINSEGLUTHTHLEASPROMETLYSVALPROASPHISALASPLYSPhEGLUATGHis 61
DB 2995 AACAAATGAA-----ATGATTAAGCATTTTATAGTAAATTTGATATGTA 3042
QY 62 VALGILYILEVALASPhELYSGLYLULEUALIAMELARGASNILEGLUALIARGGLYLEU 81
DB 3043 AAAGCAATTCAGAAATAATTTAATTAACAGATGTTTCGAAAGTATAGAACCACTATAGTA 3102
QY 82 LYSGLIMELYSARGLINGLILYASPALASNVALLYSGLYUGLUYILEVALIYSAIA 101
DB 3103 ATCCAAATTCAGAAAGAAAGGTGATTTGAAAT-----GAAATGTGTTAGTTG 3150
QY 102 HISLEULEULEGLIYVALHISASPSPILEVALSERMETGLUTYRASPLeUALIATYRYS 121
DB 3151 ATTTTA-----GATATATAGAAATAATGAAACAGTTTATTAATAATA 3195
QY 122 LEUGLYASPLEUHIProTHrHISVALIIESERASPILEGLINASPhEVALIYALIA 141
DB 3196 TTAGAAAATATTTCAAGTCTGACAGGTGTTCAAGAAACTGTACAGACATGTA----- 3249
QY 142 LEUSERLEUGLUILESERASPIGLIYASNIIETHMETHTSERHEGLUVAL----- 159
DB 3250 -----GACAAAATGTA---TATGTGAGTTGATGTTCTGCT 3285
QY 160 -----ArgGLINPhEALIASNVALLIASNHSIIleGILYLeUSER----- 173
DB 3286 ATGAAAGATCAATTTTATAGCAATATTAATGAGCGAGAGGTGTAAGAAATGTTT 3345
QY 174 ILEUASPPROIIlePhEGLIYVALLEUSERASPVALLLeUTHRALA-----IIlePhEGLIN 191
DB 3346 AATTTCGAAGATGTATTAAAGTGAAGTGAATTAATTACTGTAGAGAAATTAAGAT 3405
QY 192 ASPTHTVALARGLYSGLUETHRILYSVALLEUALIAPROALAPHELYSARGLUUGLU 211
DB 3406 GAACCGGTTCAAAAAGAGGTAGAAAAGAACTGTAGTATTATTGAA---GAATATGGA 3462
QY 212 LYSASN 213
DB 3463 GAAAT 3468

RESULT 5
US-08-973-462-1
Sequence 1, Application US/08973462B
Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DRUIHE PIERRE
APPLICANT: DAUBERSTES, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007

EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 6152
TYPE: DNA
ORGANISM: P. falciparum
US-08-973-462-1

Alignment Scores:
Pred. No.: 0.457 Length: 6152
Score: 90.00 Matches: 47
Percent Similarity: 44.55% Conservative: 43
Best Local Similarity: 23.27% Mismatches: 78
Query Match: 8.43% Indels: 34
DB: Gaps: 9

US-10-024-955-7 (1-213) x US-08-973-462-1 (1-6152)

QY 22 TYRASPPLYSILETHRGUUGLUILEASNLYSALAILLEASPSAPALAILLEALAILLEGLU 41
DB 3179 TTTTAATCTGTATTAGATTAAGTAGAGAAACGTGAAATTTAGCGAGAAATTTAGAA 3238
QY 42 GINSEGLUTHTHLEASPROMETLYSVALPROASPHISALASPLYSPhEGLUATGHis 61
DB 3239 AACAAATGAA-----ATGATTAAGCATTTTATAGTAAATTTGATATGTA 3286
QY 62 VALGILYILEVALASPhELYSGLYLULEUALIAMELARGASNILEGLUALIARGGLYLEU 81
DB 3287 AAAGCAATTCAGAAATAATTTAATTAACAGATGTTTCGAAAGTATAGAACCACTATAGTA 3346
QY 82 LYSGLIMELYSARGLINGLILYASPALASNVALLYSGLYUGLUYILEVALIYSAIA 101
DB 3347 ATCCAAATTCAGAAAGAAAGGTGATTTGAAAT-----GAAATGTGTTAGTTG 3394
QY 102 HISLEULEULEGLIYVALHISASPSPILEVALSERMETGLUTYRASPLeUALIATYRYS 121
DB 3395 ATTTTA-----GATATATAGAAATAATGAAAGAGTTTATTAATAATA 3439
QY 122 LEUGLYASPLEUHIProTHrHISVALIIESERASPILEGLINASPhEVALIYALIA 141
DB 3440 TTAGAAAATATTTCAAGTCTGACAGGTGTTCAAGAAACTGTACAGACATGTA----- 3493
QY 142 LEUSERLEUGLUILESERASPIGLIYASNIIETHMETHTSERHEGLUVAL----- 159
DB 3494 -----GACAAAATGTA---TATGTGAGTTGATGTTCTGCT 3529
QY 160 -----ArgGLINPhEALIASNVALLIASNHSIIleGILYLeUSER----- 173
DB 3530 ATGAAAGATCAATTTTATAGCAATATTAATGAGCGAGAGGTGTAAGAAATGTTT 3589
QY 174 ILEUASPPROIIlePhEGLIYVALLEUSERASPVALLLeUTHRALA-----IIlePhEGLIN 191
DB 3590 AATTTCGAAGATGTATTAAAGTGAAGTGAATTAATTACTGTAGAGAAATTAAGAT 3649
QY 192 ASPTHTVALARGLYSGLUETHRILYSVALLEUALIAPROALAPHELYSARGLUUGLU 211
DB 3650 GAACCGGTTCAAAAAGAGGTAGAAAAGAACTGTAGTATTATTGAA---GAATATGGA 3706
QY 212 LYSASN 213
DB 3707 GAAAT 3712

RESULT 6
US-08-961-527-79/C
Sequence 79, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.

|||||
Db 368 GGTCAATTGTTGGTCGGTGTTCATGATGACGTTGTTCAATGCAATATGATTAGCATAC 427
Qy 121 LysLeuGlyAspLeuHisProThrRHisValIleSerAspIleGlnAspPheVal 140
Db 428 AATTGGGATGCTTCATCCAAACACTCATGTCATTTCCGATTTAGCATTTTGTGTGC 487
Qy 141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrsSerPheGluValArg 160
Db 488 GAATTAATCCCTCGAAGTATGAGAGAAAGTAATATGACATTCGATTCGAAATGACGT 547
Qy 161 GlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
Db 548 CAATTGGCCAATGTTGTCATCATATTGGTGTCTTCAATTTTGGATCCAAATTTTCGCT 607
Qy 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgGlyGluMetThrs 200
Db 608 GTCTTATCCGATGTTTGGACCGCTATTTCAGAGATACCGTACGTCAGAAATGACCAAA 667
Qy 201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
Db 668 GTATTGGCACACGATTCAAAAAAGAAATTGGAACGAAC 706

RESULT 3
US-08-462-778-1
; Sequence 1, Application US/08462778
; Patent No. 6077517
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: China, Kaw-Tan
; TITLE OF INVENTION: Allergenic Protein and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Labive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,778
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/031,141
; FILING DATE: 12 March 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 812 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 68..712
; US-08-462-778-1

Alignment Scores: 1.85e-112 Length: 812
Pred. No.: 1.85e-112

Score: 959.00 Matches: 182
Percent Similarity: 94.37%
Best Local Similarity: 85.45%
Query Match: 89.79%
DB: 3 Gaps: 0

US-10-024-955-7 (1-213) x US-08-462-778-1 (1-812)

Qy 1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
Db 68 ATGATCAAAATTAATTAATGATGCTGCGCAGCGCTTTGTGTCGCTTGGCTATCAATT 127
Qy 21 HisTyrAspLysIleThrGluGluLeuAsnLysAlaIleAspAlaIleAlaIle 40
Db 128 CACTATGATAAATACACGAGAAATTAACAAGCTGTGATGAAGACCGCTCGCAATT 187
Qy 41 GluGlnSerGluThrIleAspPheMetLysValProAspHisAlaAspLysPheGluArg 60
Db 188 GAAAAATCCGAAACATTCGATCCAAATGAAGATACCGATCATTCGATTAATTCGAACGA 247
Qy 61 HisValGlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGly 80
Db 248 CATATTGGTATCATCATTTAAAGCTCAATTAGACATCGAAACATTCAGTTCGAGCA 307
Qy 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGluGluGlyIleValLys 100
Db 308 TTAANAACAAATGAACGCTAGATGATGCTAATGTGAAGAAGATGATGCTGTTCANA 367
Qy 101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
Db 368 GGTCAATTGTTGGTCGGTGTTCATGATGACGTTGGTTCATGAATATGATTTAGCATAC 427
Qy 121 LysLeuGlyAspLeuHisProThrRHisValIleSerAspIleGlnAspPheValVal 140
Db 428 AATTGGGATGCTTCATCCAAACACTCATGTCATTTCCGATTTAGCATTTTGTGTGC 487
Qy 141 AlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrsSerPheGluValArg 160
Db 488 GAATTAATCCCTCGAAGTATGAGAGAAAGTAATATGACATTCGATTCGAAATGACGT 547
Qy 161 GlnPheAlaAsnValValAsnHisIleGlyGlyLeuSerIleLeuAspProIlePheGly 180
Db 548 CAATTGGCCAATGTTGTCATCATATTGGTGTCTTCAATTTTGGATCCAAATTTTCGCT 607
Qy 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgGlyGluMetThrs 200
Db 608 GTCTTATCCGATGTTTGGACCGCTATTTCAGAGATACCGTACGTCAGAAATGACCAAA 667
Qy 201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
Db 668 GTATTGGCACACGATTCAAAAAAGAAATTGGAACGAAC 706

RESULT 4
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILLHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum


```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 761 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 43..681
: US-08-553-336A-6

Alignment Scores:
Pred. No.: 2.58e-126 Length: 761
Score: 1068.00 Matches: 213
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-08-553-336A-6 (1-761)
QY 1 MetMetLysPheLeuLeuIleAlaIaValAlaIaPheValAlaIaSerLysAlaProIle 20
DB 43 ATGATGAATATTTTGTGATTCCTCCGCGCATTTGTCGCCGTTTCGGCTGATCCAATT 102
QY 21 HistYrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAlaIaIle 40
DB 103 CACTATGATTAATCAACCGAACAATCAACAAGCTATTGATGCCATTCCTCTCTATT 162
QY 41 GluInSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGluArg 60
DB 163 GAACAATCCGAACAATAGATTCATGAAGAATCACTGATCATCGCATTAATTCGAACGT 222
QY 61 HisValGlyIleValAspPheLysGlyIleuAlaIleMetArgAsnIleGluAlaArgGly 80
DB 223 CATGTTGGATGTGATTCGAAGTTCAGAGTATGACCATCGCAACATTTGAGCCTGAGGA 282
QY 81 LeuLysGlnMetLysArgGlnGlnLysAspAlaAsnValLysGlyGluGluIleValLys 100
DB 283 TTGAACAACAATGAACGTCACAGTGATGCTAATGTCACAAGGAGAGGATTTGTTAA 342
QY 101 AlaHisLeuLeuIleGlyIleHisAspAspIleValSerMetGluThrAspLeuAlaTyr 120
DB 343 GCTCATTTGTTGATCGGTGTTCACGATGATATCGTCTCGATGAATATGATTTAGCATAC 402
QY 121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValVal 140
DB 403 AATTGGGTGATCTTCATCCCAACCACTCATGTCATTTCCGATATTCACAGATTTGTTGTT 462
QY 141 AlaLeuSerLeuGluIleSerAspGlnGluIleAsnIleThrMetThrSerPheGluValArg 160
DB 463 GCGTTGTCCTTGAATATTTCTGATGAGGTACATACATACATACATCTTTGAAGTACGA 522
QY 161 GlnPheAlaAsnValValAlaHisIleGlyIleLeuSerIleLeuAspProIlePheGly 180
DB 523 CATATTCGTAAGTGTGTCAACCATATTTGGTGTCTTCAATCTTGATCCAAATTTTGGC 582
QY 181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys 200
DB 583 GTTTATATCGATGATGACCGCTATTTTCCAAAGACACCGTCGTAAGAAATGACCAAA 642
QY 201 ValLeuAlaProAlaPheLysArgGluLeuGluLysAsn 213
DB 643 GATATGGCACCGACGATTTAAACGTGAAATTTGAAAAAAT 681

RESULT 2
: Sequence 1, Application US/08553336A
: Patent No. 6413738
: GENERAL INFORMATION:
: APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
: TITLE OF INVENTION: Allergenic Proteins and Peptides From
: TITLE OF INVENTION: House Dust Mite and Uses Therefor

```

```

: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD, LLP
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/553,336A
: FILING DATE: 10-JUN-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/081,540
: FILING DATE: 22-JUNE-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Jane E. Remillard
: REGISTRATION NUMBER: 38,872
: REFERENCE/DOCKET NUMBER: IMI-032CP2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)742-4214
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 812 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 68..712
: NAME/KEY: mat_peptide
: LOCATION: 119..712
: US-08-553-336A-1

Alignment Scores:
Pred. No.: 7.69e-113 Length: 812
Score: 962.00 Matches: 183
Percent Similarity: 94.37% Conservative: 18
Best Local Similarity: 85.92% Mismatches: 12
Query Match: 90.07% Indels: 0
DB: 4 Gaps: 0

US-10-024-955-7 (1-213) x US-08-553-336A-1 (1-812)
QY 1 MetMetLysPheLeuLeuIleAlaIaValAlaIaPheValAlaIaSerLysAlaProIle 20
DB 68 ATGATGAATATTTTGTGATTCCTCCGCGCATTTGTTGTCGCCGTTTCGGCTGATCCAANT 127
QY 21 HistYrAspLysIleThrGluGluIleAsnLysAlaIleAspAspAlaIleAlaIaIle 40
DB 128 CACTATGATTAATCAACCGAACAATTAACAAGCTGTTGATGAAGCCGTCCTCAANT 187
QY 41 GluInSerGluThrIleAspProMetLysValProAspHisAlaAspLysPheGluArg 60
DB 188 GAANAATCCGAACAATTCGATTCATGAAGTACCCGATCATTCGATTAATTCGAACGA 247
QY 61 HisValGlyIleValAspPheLysGlyIleuAlaIleMetArgAsnIleGluAlaArgGly 80
DB 248 CATATTTGATCATCGATTTAAAGTGAATTAAGTCAATGACATCCGATCAATTAATTCGAACGA 307
QY 81 LeuLysGlnMetLysArgGlnGlnLysAspAlaAsnValLysGlyGluGluIleValLys 100
DB 308 TTAAAAACAATGAACGTCGTAGTGTATGATGATGTAAGAAAGAGAGATGTTGTCAAA 367
QY 101 AlaHisLeuLeuIleGlyIleHisAspAspIleValSerMetGluThrAspLeuAlaTyr 120

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 16:56:36 ; Search time 78 Seconds
(without alignments)
837.463 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 1068
Sequence: 1 MKKFLIAVAFAVAVSADPI.....VRKEMTKVLAFAKRELEKN 213

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPRO.spool/US10024955/runat_14022003_130658_23735/app_query.fasta_1.391
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=50
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USER=US10024955 @CGN.1.1.61 @runat_14022003_130658_23735 -NCPu=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6C.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfill1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1068	100.0	761	4	US-08-553-336A-6
2	962	90.1	812	4	US-08-553-336A-1
3	959	89.8	812	3	US-08-462-778-1
4	90	8.4	5361	4	US-08-973-462-2
5	90	8.4	6152	4	US-08-973-462-1
6	89	8.3	5066	4	US-08-961-527-79
7	88.5	8.3	2157	1	US-08-336-618-25
8	87.5	8.2	1155	1	US-08-002-024B-7
9	86	8.1	4411529	4	US-09-103-840A-1
10	85.5	8.0	1693	3	US-09-320-878-23
11	84.5	7.9	1958	1	US-08-137-175A-7
12	84.5	7.9	1958	3	US-08-479-017-7

13	84.5	7.9	2472	4	US-09-134-001C-1244	Sequence 1244, Ap
14	82	7.7	1089	4	US-09-134-001C-1220	Sequence 1220, Ap
15	81.5	7.6	1647	4	US-09-472-971-5	Sequence 7, Appl1
16	81.5	7.6	4524	4	US-09-472-971-7	Sequence 7, Appl1
17	79.5	7.4	3123	1	US-08-094-889-2	Sequence 2, Appl1
18	79	7.4	1365	4	US-09-450-433-1	Sequence 1, Appl1
19	78.5	7.4	2019	3	US-09-040-843-3	Sequence 3, Appl1
20	78.5	7.4	2019	4	US-09-621-855-3	Sequence 3, Appl1
21	78.5	7.4	2599	3	US-09-040-843-1	Sequence 1, Appl1
22	78.5	7.4	2599	3	US-09-621-855-1	Sequence 1, Appl1
23	78.5	7.4	19227	3	US-09-090-793-13	Sequence 13, Appl1
24	78.5	7.4	40138	3	US-09-090-793-12	Sequence 12, Appl1
25	78	7.3	1362	4	US-09-232-479-13	Sequence 13, Appl1
26	78	7.3	1362	4	US-09-784-990-13	Sequence 13, Appl1
27	78	7.3	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
28	78	7.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
29	77.5	7.3	4443	4	US-09-425-453A-7	Sequence 7, Appl1
30	77.5	7.3	4443	4	US-09-425-453A-17	Sequence 17, Appl1
31	77	7.2	3261	2	US-08-708-541A-27	Sequence 27, Appl1
32	77	7.2	3261	2	US-08-708-541A-29	Sequence 29, Appl1
33	77	7.2	3261	4	US-08-940-968-23	Sequence 23, Appl1
34	76.5	7.2	2378	4	US-09-521-780-1	Sequence 1, Appl1
35	76	7.1	3588	1	US-08-307-382-1	Sequence 1, Appl1
36	76	7.1	3588	1	US-08-366-779-1	Sequence 1, Appl1
37	76	7.1	3588	1	US-08-478-727-1	Sequence 1, Appl1
38	76	7.1	3588	1	US-08-473-508-1	Sequence 1, Appl1
39	76	7.1	3588	1	US-08-769-936-1	Sequence 1, Appl1
40	76	7.1	3588	4	US-08-934-254-1	Sequence 1, Appl1
41	76	7.1	9069	4	US-08-961-527-97	Sequence 97, Appl1
42	75.5	7.1	1471	5	PCT-US91-03056-15	Sequence 15, Appl1
43	75.5	7.1	3180	1	US-07-944-943-1	Sequence 1, Appl1
44	75.5	7.1	3180	1	US-07-944-525-1	Sequence 1, Appl1
45	75.5	7.1	3180	1	US-08-219-262B-13	Sequence 13, Appl1

ALIGNMENTS

RESULT 1
US-08-553-336A-6
Sequence 6, Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
NUMBER OF INVENTIONS: House Dust Mite and Uses Therefor
CORRESPONDENCE ADDRESS: 15
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-0320CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 6:


```
Db 2796 GCCGGCCTGACACCCCTCTAGATTCTATTCCACGCTATGGGTGT---CGTGAAGGTCTT 2852
QY 140 ValAlaLeuSerLeuGluIleSerAspGluGlyAsnIle-----Thr 153
   ::::::::::: ::::: |||
Db 2853 ATGGATACGCTGTATAAGACAGCTGAACCGTTACATCCAGCGTGTCTTATAAGGCT 2912
QY 154 MetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeuSer 173
   ||| ||| ||| ||| ||| ||| ::::: |||
Db 2913 ATGAGTCCGTTGATGTTAACTACGACGAGAACAGTGCCTAACTCGGTGGGCCAGCTTATT 2972
QY 174 IleLeu-----AspProIlePheGlyValLeuSerasp----- 184
   ||| ||| ::::: ||| ||| :::::
Db 2973 CAGCTGCGTTACGGCGAGGACGGCTCTTGCGGGAGCTGTTGAGTTCAGAACATGCCA 3032
QY 184 ----- 184
Db 3033 ACGGTGAACCTGTCGAACAAGTCGTTTGAAAAGCGCTTCAATTTCAGCTGGAGCAACGAG 3092
QY 185 ---ValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThr 199
   ::::: ::::: ||| ||| ||| ||| |||
Db 3093 CGATTATGATAAAGSTATTACGATGATGATCATCAGAGATGAC 3140
```

Search completed: February 20, 2003, 16:53:25
Job time : 1605 secs

DB 1448 CCTAGCATATTATATCTTCTACAGTTTGTAGATGATGCTATTTACAGATACTATAA 1389
QY 157 PheGluValAlaArgInPheAlaAsnValVal-----AsnHisIleGlyGlyLeu 172
DB 1388 GGTGCGACTGTGACCTTCCGTAATCTGATATTATTTAGACTTGAATGGGAGCTCA 1329
QY 173 SerIleuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp 192
DB 1328 GAATTTACAGGACCAACGCTTTGCTGCTTTGGAGCTCTTCAGAGCTAGTACTACGAA 1269
QY 193 ThrValAlaArgIleGluMetThrLysValLeuAlaProAlaPheLysArgGlu 209
DB 1268 ACTGTCAGAAACAAATGATGAAAGAAATTTCAATTCGACGACGAGAA 1218
RESULT 47
AAH54312/C
ID AAH54312 standard; DNA: 3295 BP.
XX
AC AAH54312;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3676.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX
KM vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000MO-US30782.
XX
PR 09-NOV-1999; 9905-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1285-1287; 2188bp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG81320, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SO Sequence 3295 BP; 971 A; 667 C; 456 G; 1201 T; 0 other;

Alignment Scores:

Pred. No.:

10.3

Length:

3295

Score: 84.50 Matches: 50
Percent Similarity: 39.17% Conservative: 35
Best Local Similarity: 23.04% Mismatches: 87
Query Match: 7.91% Indels: 45
DB: 22 Gaps: 11
US-10-024-955-7 (1-213) x AAH54312 (1-3295)
QY 18 AspProIleHis-----TyrAspLysIleThrGluIleAsnLysAla 32
DB 1957 GATACACTTCAATAACGTCGATTCGACAAACAGAGCTGTCAATTCATTAAGTAAAGCT 1898
QY 33 IleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspPrometLysValPro 52
DB 1897 GTTAGAAGAGCTCGTGTGCTTTAA-----GATCA----- 1865
QY 53 AspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyLeuAla 72
DB 1864 -----AAACGTCMAATCGTAGTTTATTTCTTAGACGCTACAGCT 1823
QY 73 MetArgAsnIleGlu---AlaArgGlyLeu---LysGlnMetLysArgGlnGlyAspAla 90
DB 1822 GTGGGTAACCTGAATTTGGCTGCTTGTAGCTGAATCTATGTTGGTGAAGACGATGA 1763
QY 91 AsnValLysGlyGlu---GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
DB 1762 ATGATTCGCGCTAGATATGATGATGAATTTATGACAAACATCTGTCAGTCATTAAGTTGCT 1703
QY 107 Val-----HisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
DB 1702 GCACCTTCACAGATATGTAGGACATGATGACGCGCTCAATTTGACTGMAAAAGTTAGACGT 1643
QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAsp-----IleGln 136
DB 1642 AAA-----CATACTCTGTGATTTTATTTATGTAATGCAAAATGCAAAACGACAT 1598
QY 137 AspPheValAlaLeuSerLeuGluIleSerAspGluLysAsnIleThrMetThrSer 156
DB 1597 CTTGACGTATTTAATATCTCTACAAAGTTTATGATGATGTCATTAAACGATACTATAA 1538
QY 157 PheGluValAlaArgInPheAlaAsnValVal-----AsnHisIleGlyGlyLeu 172
DB 1537 GGTGCGACTGTGACCTTCCGTAATCTGATATTATTTAGACTTCTAATGTGCGAGCTCAA 1478
QY 173 SerIleuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp 192
DB 1477 GAATTTACAGGACCAACGCTTTGCTGCTTTGGAGCTGCTTCAGAGCTAGTACTACGAA 1418
QY 193 ThrValAlaArgIleGluMetThrLysValLeuAlaProAlaPheLysArgGlu 209
DB 1417 ACTGTCAGAAACAAATGATGAAAGAAATTTCAATTCGACGACGAGAA 1367
RESULT 48
AAH54712/C
ID AAH54712 standard; DNA: 3742 BP.
XX
AC AAH54712;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4076.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
KM vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000MO-US30782.
XX

```

Query Match: 7.91% Indels: 45
DB: 24 Gaps: 11
US-10-024-955-7 (1-213) x ABN91781 (1-2472)

QY 18 AspProIleHis-----TyrAspLysIleThrGluGluIleAsnLysAla 32
   ||| :|||
   ||| :|||
DB 1528 GATACACTTCATTAACGTCGTCATTTGAGCAAAACATGCTGTCATTCATTAAGCT 1587
   ||| |||
   ||| |||
QY 33 IleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspProMetLysValPro 52
   ||| |||
   ||| |||
DB 1588 GTTAGAAGACCTGCTGCTGCTTAA-----GATCCA----- 1620
   ||| |||
   ||| |||
QY 53 AspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyIleValAla 72
   ||| |||
   ||| |||
DB 1621 -----AAAGCTCCAAATCGAGTATTTATTTCTTAGACCTACAGCT 1662
   ||| |||
   ||| |||
QY 73 MetArgAsnIleGlu--AlaArgGlyLeu--LysGlnMetLysArgGlnGlyAspAla 90
   ||| |||
   ||| |||
DB 1663 GTGGGTAAACCTGAATTGGCTGCTGCTGCTGTAATCTGTATTTGGTGAAGACGATGCA 1722
   ||| |||
   ||| |||
QY 91 AsnValLysGlyLeu--GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
   ||| |||
   ||| |||
DB 1723 ATGATTGCGTAGATATGAGTGAATTTATGAGAAACATGCTCTCAGTCGATTAGTTGGT 1782
   ||| |||
   ||| |||
QY 107 Val-----HisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
   ||| |||
   ||| |||
DB 1783 GCACCTCCAGATATGTAGACATGATGACGGGCGTCATTCGTAAGTTAGACGT 1842
   ||| |||
   ||| |||
QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAsp-----IleGln 136
   ||| |||
   ||| |||
DB 1843 AAA-----CATCTACTGCTGATTTTATTTGTAAGTAATTTGAGAAAGCACAT 1887
   ||| |||
   ||| |||
QY 137 AspPheValValAlaLeuSerLeuGluIleSerAspGlnLysAsnIleThrMetHisSer 156
   ||| |||
   ||| |||
DB 1888 CCTGACGATATTTATATCTTCTTACAGATTGATGATGATGATGATTAACAGTACTAAA 1947
   ||| |||
   ||| |||
QY 157 PheGluValArgGlnPheAlaAsnValVal-----AsnHisIleGlyGlyLeu 172
   ||| |||
   ||| |||
DB 1948 GGTGCTACTGTGACCTCCGTATACGTGATTTATGACTCTTATATGCTTATATGCGGAGCTCAA 2007
   ||| |||
   ||| |||
QY 173 SerIleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp 192
   ||| |||
   ||| |||
DB 2008 GAATTCACGACGACACGCTTCTGCTTTGAGAGTGCCTTCAGAAAGTAGTACCTACGAA 2067
   ||| |||
   ||| |||
QY 193 ThrValArgLysGlnMetThrLysValLeuAlaProAlaPheLysArgGlu 209
   ||| |||
   ||| |||
DB 2068 ACTGTCAAGAAACATGATGATAAATTTAAATTCATTCGACCCAGAA 2118
   ||| |||
   ||| |||
RESULT 46
AAH54604/c
AAH54604 standard; DNA; 3002 BP.
XX AC AAH54604;
XX AC
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3968.
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000MO-US30782.
XX PR 09-NOV-1999; 9905-0164258.
XX PA (GLAXO ) GLAXO GROUP LTD.

```

```

XX XX
PI kimmerly WJ;
XX MPI; 2001-316495/33.
DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 8; Page 1634-1635; 2186pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAH51454 to AAH51120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to
CC AAH5098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 3002 BP; 899 A; 595 C; 418 G; 1090 T; 0 other;
XX
Alignment Scores:
Pred. No.: 9.1 Length: 3002
Score: 84.50 Matches: 50
Percent Similarity: 39.17% Conservative: 35
Best Local Similarity: 23.04% Mismatches: 87
Query Match: 7.91% Indels: 45
DB: 22 Gaps: 11
US-10-024-955-7 (1-213) x AAH54604 (1-3002)

QY 18 AspProIleHis-----TyrAspLysIleThrGluGluIleAsnLysAla 32
   ||| :|||
   ||| :|||
DB 1808 GATACACTTCATTAACGTCGTCATTTGAGCAAAACATGCTGTCATTCATTAAGCT 1749
   ||| |||
   ||| |||
QY 33 IleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspProMetLysValPro 52
   ||| |||
   ||| |||
DB 1748 GTTAGAAGACCTGCTGCTGCTTAA-----GATCCA----- 1716
   ||| |||
   ||| |||
QY 53 AspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyIleValAla 72
   ||| |||
   ||| |||
DB 1715 -----AAAGCTCCAAATCGAGTATTTATTTCTTAGACCTACAGCT 1674
   ||| |||
   ||| |||
QY 73 MetArgAsnIleGlu--AlaArgGlyLeu--LysGlnMetLysArgGlnGlyAspAla 90
   ||| |||
   ||| |||
DB 1673 GTGGGTAAACCTGAATTGGCTGCTGCTGCTGTAATCTGTATTTGGTGAAGACGATGCA 1614
   ||| |||
   ||| |||
QY 91 AsnValLysGlyLeu--GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
   ||| |||
   ||| |||
DB 1613 ATGATTGCGTAGATATGAGTGAATTTATGAGAAACATGCTGTCAGTCGATTAGTTGGT 1554
   ||| |||
   ||| |||
QY 107 Val-----HisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
   ||| |||
   ||| |||
DB 1553 GCACCTCCAGATATGTAGACATGATGACGGGCGTCATTCGTAAGTAAGTTAGACGT 1494
   ||| |||
   ||| |||
QY 121 LysLeuGlyAspLeuHisProThrHisValIleSerAsp-----IleGln 136
   ||| |||
   ||| |||
DB 1493 AAA-----CATCTACTGCTGATTTTATTTATGATAAATTCGAAAGCACAT 1449
   ||| |||
   ||| |||
QY 137 AspPheValValAlaLeuSerLeuGluIleSerAspGlnLysAsnIleThrMetHisSer 156
   ||| |||
   ||| |||

```



```

PR 09-NOV-1999; 990S-0164258.
XX
XX (GLAXO ) GLAXO GROUP LTD.
XX
XX kimmerly wj;
PI
XX WPI: 2001-316495/33.
DR P-PSDB: AAG82217.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; page 431-432; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 2454 BP; 895 A; 349 C; 514 G; 696 T; 0 other;

Alignment Scores:
Pred. No.: 6.91 Length: 2454
Score: 84.50 Matches: 50
Percent Similarity: 39.17% Conservative: 35
Best Local Similarity: 23.04% Mismatches: 87
Query Match: 7.91% Indels: 45
DB: Gaps: 11

US-10-024-955-7 (1-213) x AAH53067 (1-2454)
OY 18 AspProIleHis-----TyrAspLysIleThcGluGluIleAsnLysAla 32
DB 1510 GATACACTTCATTAACGTCGTGATGACAAACGATGCTGCAATTCATTTAGTAACGCT 1569
OY 33 IleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspProMetLysValPro 52
DB 1570 GTTAGAAGAGCTGCTGCTGCTTAA-----GATCCA-----1602
OY 53 AspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGlyLueAla 72
DB 1603 -----AAAGCTCAACGGAGTTATTATTATTCTTAGACCTACAGCT 1644
OY 73 MetArgAsnIleGlu---AlaArgGlyLeu---LysGlnMetLysArgGlnGlyAspAla 90
DB 1645 GTGGGTAAACATGAATGGCTGCTGCTTAGCTGAATGATATGTTGGTAAAGCATGCA 1704
OY 91 AsnValLysGlyGlu---GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
DB 1705 ATGATTCGGGTAGATATGATGCAATTTATGGAACACGCTGCTCAGTGGATTAGTTGGT 1764
OY 107 Val-----HisAspAspIleValSerMetClnuThrLysAspLeuAlaTyr 120
DB 1765 GCACCTCCAGATATGTAGACATGATGACGCGGTCAATTCAGTCAAAAGTTAGACGT 1824
OY 121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAsp-----IleGln 136

```

```

DB 1825 AAA-----CCATACCTGTGATTTTATTGTAAGAAATTGAGAAACACAT 1869
OY 137 AspPheValIaIaLeuSerLeuGluIleSerAspGluGlyAsnIleTherThrSer 156
DB 1870 CCTGACGTATTAAATATCTTCTTCAAGTTTAAAGATGATGATTAACGATACTAA 1929
OY 157 PheGluValArgGlnPheAlaAsnValVal}-----AsnHisIleGlyGlyLeu 172
DB 1930 GGTGCTAGTGTGACTTCCTGCTAATGATGATTAATGATGATTCATGATGGAGCTCAA 1989
OY 173 SerIleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAsp 192
DB 1990 GAATTAACAGGACCAACGCTTGTGCTGTTTGGAGCTGCTTCAGAAAGTACGACTACGAA 2049
OY 193 ThrValArgLysGluMetThrLysValLeuAlaProAlaPheLysArgGlu 209
DB 2050 ACTGTCAGAAAAACAATGATGTAAGAATTAATAAAATTCTTCGACACGAA 2100

RESULT 45
ABN91781
ID ABN91781 standard; DNA; 2472 BP.
XX
XX ABN91781;
AC
XX 24-JUL-2002 (first entry)
DT
XX
XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1244.
DE
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy; gene; ds.
XX
XX Staphylococcus epidermidis.
PN
XX US6380370-B1.
PD
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
PF
XX
XX 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX
XX Doucette-Stamm LA, Bush D;
PI
XX
XX WPI: 2002-381255/41.
DR P-PSDB: ABP39236.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 1244; 267pp; English.
PS
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 2472 BP; 903 A; 351 C; 518 G; 700 T; 0 other;

Alignment Scores:
Pred. No.: 6.98 Length: 2472
Score: 84.50 Matches: 50
Percent Similarity: 39.17% Conservative: 35
Best Local Similarity: 23.04% Mismatches: 87

```

```
Qy 173 -----SerIleuAspProIlePheGlyValIleuSerAsp----- 184
Cc      |||||  |||  |||||:::  |||
Db 2387 CAGCGATTTCACCAAGATCTGACTTAGATTGAAGCGGTGTGCTGACCCGACGCTCCAG 2328
Qy 185 ValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLysValIleuAlaPro 204
Cc      ::::  ::::|  |||  :::::  ::::|  |||||
Db 2327 CTGACTGTGCAGCTGTGGAAAGCGCTGTGATGACACACGTCATCCGCTGGCGGCGCG 2268
Qy 205 AlaphelysArgLuleu 210
Cc      ::::  |||
Db 2267 GGAGTAGAAACTTCTCTC 2250

RESULT 43
AAH52328
ID AAH52328 standard; DNA; 2454 BP.
Cc
AC AAH52328;
Cc
DT 03-SEP-2001 (first entry)
Cc
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:49.
Cc
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
Km vaccination; endocarditis; ds.
Cc
XX Staphylococcus epidermidis.
Cc
PN WO200134809-A2.
Cc
PD 17-MAY-2001.
Cc
PF 09-NOV-2000; 2000WO-US30782.
Cc
PR 09-NOV-1999; 99US-0164258.
Cc
XX (GLAXO ) GLAXO GROUP LTD.
Pa
PI Kimerly WJ;
Cc
DR WPI: 2001-316495/33.
Cc
XX P-PSDB: AAG81478.
Cc
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
Cc
XX useful for vaccinating against infections, e.g. endocarditis -
Cc
PS Claim 8; Page 58-59; 2188pp; English.
Cc
XX
Cc AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
Cc (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
Cc (I) and (II) can have antibacterial activity and therefore can be used
Cc in vaccination. The nucleic acids (I) may be used to produce the
Cc S. epidermidis polypeptides (II) via the production of vectors
Cc containing them which are used to produce hosts cells which express the
Cc polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
Cc used to vaccinate subjects and to raise antibodies against the bacteria.
Cc The polypeptides may also be used to assay for other inhibitors of their
Cc activity and therefore identify compounds that may be used for the
Cc treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
Cc AAH55090 represent specifically claimed S. epidermidis genomic DNA
Cc polynucleotide sequences from the present invention. AAH55091 to
Cc AAH55098 represent oligonucleotide sequences and primers which are used
Cc in the exemplification of the present invention.
Cc N.B. The present invention specifically claims all the polynucleotide
Cc sequences given in the sequence listing of the present specification,
Cc however the sequence listing only goes up to SEQ ID NO:4454 so even
Cc though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
Cc no sequences are present for SEQ ID NO:4455 to 4464.
Cc
XX
SQ Sequence 2454 BP; 895 A; 349 C; 514 G; 696 T; 0 other;

Alignment Scores: 6.91 Length: 2454
Pred. No.: 
```

```
Score: 84.50 Matches: 50
Percent Similarity: 39.17% Conservative: 35
Best Local Similarity: 23.04% Mismatches: 87
Query Match: 7.91% Indels: 45
DB: 22 Gaps: 11

US-10-024-955-7 (1-213) x AAH52328 (1-2454)
Qy 18 AsproIleHis-----TyrAspLysIleThrGluGluIleAsnLysAla 32
Cc      |||  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
Db 1510 GATACACTTCATAAACGTCGTCTATTGGACAAACGATGCTGTCATATTCATTAAGTAACCT 1569
Qy 33 IleAspAspAlaIleAlaAlaIleGluGlnSerGluThrIleAspPrometLysValPro 52
Cc      ::::  |||  |||  :::::  ::::|  |||||
Db 1570 GTTACAGAGAGCTGCTGCTGTCTTAAA-----GATCCA----- 1602
Qy 53 AspHisAlaAspLysPheGlnArgHisValGlyIleValAspPheLysGlyLuleuAla 72
Cc      ::::|  ::::|  ::::|  ::::|  |||  |||  |||
Db 1603 -----AAACGTCCAAATCGGTACTTTTATTTTCTTAGAGACTTACAGCT 1644
Qy 73 MetArgAsnIleGlu--AlaArgGlyLeu--LysGlnMetLysArgGlnGlyAspAla 90
Cc      ::::  ::::|  |||  |||||  |||  ::::  |||  ::::  |||||
Db 1645 GTGGGTAAACATGAATTGGCTGCTGCTTACGTGATCATGTGTTGGTGAACAGCATGCA 1704
Qy 91 AsnValLysGlyGlu--GluGlyIleValLysAlaHisLeu-----LeuIleGly 106
Cc      :::::  ::::  :::::  |||  :::::  |||  |||  ::::|  |||
Db 1705 ATGATTCCGCTAGATATGATGAATTTATGAGAAACATGCTGCATGATTAAGTTGCT 1764
Qy 107 Val-----HisAspAspIleValSerMetGluThrAspLeuAlaIyr 120
Cc      ::::|  |||||  |||  ::::|  |||  ::::|  |||  ::::|  |||
Db 1765 GCACCTCCAGATATATGACATGATGACGGCGGTCAATTGACTGAAAGAAAGTTAGACGT 1824
Qy 121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAsp-----IleGln 136
Cc      |||  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1825 AAA-----CCATACCTCTGATTTTATTTGATGAATTCAGAAACACACAT 1869
Qy 137 AspPheValAlaIleuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSer 156
Cc      |||  ::::  ::::  |||  ::::  |||  ::::  |||  ::::  |||
Db 1870 CCTGACGTATTAAATATCTCTTACACAGTTTATGATGATGATTAACATACATCTAAA 1929
Qy 157 PheGluValArgGlnPheAlaAsnValAla-----AsnHisIleGlyGlyLeu 172
Cc      |||  |||  |||  |||  ::::  |||  ::::  |||  ::::  |||
Db 1930 GGTCTACTGTGACCTCCGATATCTGATTTATGACTTCTTAATGTGGACCTCAA 1989
Qy 173 SerIleuAspProIlePheGlyValIleuSerAspValIleuThrAlaIlePheGlnAsp 192
Cc      ::::  |||  |||  ::::  |||  ::::  |||  ::::  |||
Db 1990 GAATTACAGACCAACGCTTCTGCTTTGGAGGCGCTTCAGAGGTGACTACAGCA 2049
Qy 193 ThrValArgLysGluMetThrLysValIleuAlaProAlaPheLysArgGlu 209
Cc      |||||  |||  |||  |||  ::::|  |||  ::::|  |||
Db 2050 ACTGTCAGAAAACAATGATGAAGAAATTAATAAATTCATTCGACACAGAA 2100

RESULT 44
AAH53067
ID AAH53067 standard; DNA; 2454 BP.
Cc
AC AAH53067;
Cc
DT 03-SEP-2001 (first entry)
Cc
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1527.
Cc
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
Km vaccination; endocarditis; ds.
Cc
XX Staphylococcus epidermidis.
Cc
PN WO200134809-A2.
Cc
PD 17-MAY-2001.
Cc
PF 09-NOV-2000; 2000WO-US30782.
Cc
XX 
```

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3277 BP; 824 A; 769 C; 723 G; 961 T; 0 other;

Alignment Scores:

Pred. No.:	8.94	Length:	3277
Score:	85.00	Matches:	45
Percent Similarity:	46.04%	Conservative:	48
Best Local Similarity:	22.28%	Mismatches:	85
Query Match:	7.96%	Indels:	25
DB:	23	Gaps:	8

US-10-024-955-7 (1-213) x ABL15310 (1-3277)

```

OY 21 HistyrAspLysIleThrGluIleAsnLysAlaIleAspAlaIleAlaIle 40
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1908 CACATTGCCAGCAATGCCGCCAGGTGGCCAGAGCCCAATGATGACAGCGGCCGCC 1849
OY 41 GluInserGluThrIleAsp---PrometLysValProAspHisAlaAspLysPheGlu 59
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1848 GCTCAAGATGTCATGCCGAGCGAGCAAGATGACGTGGCGAAGGCCATCTCCGCTGCC 1789
OY 60 ArgHisValGlyIleValAspPheLysGlyIleuAlaMetArgAsnIleGluAlaArg 79
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1788 CCGTCCGCCGATGCGGTG---CTCGAGGCGCAAGCGCAATGTGACAACTACGCTAGG 1732
OY 80 GlyLeuLysGlnMetLys-----ArgGlnGlyAsp 89
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1731 GAGATCCCGGATGCCGAGAGAGTGTGCGACAGGTGAGTGTGCTGCGCAACAGTGGAG 1672
OY 90 AlaAsnValLysGlyGluGlnGluIleValLysAlaHisLeuLeuIle-----GlyVal 107
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1671 ACGAACGTGAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1613
OY 108 HisAspAspIleValSerMetGluIleValAspLeuAlaIleValLysLeuGlyAspLeuHisPro 127
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1612 CAGGTATATATTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1562
OY 128 ThrThrHisValIleSerAspIleGlnAspPheValAlaIleSerLeuGluIleSer 147
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1561 ACTATCTACACATTTTAAAGCGGACGACATTAATGCTT---ATATCTATATCAATATCT 1505
OY 148 AspGluGlyAsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsn 167
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1504 ACCAGCGAATACATATCATTTTAAATATCATATGCTTTTGCATTAATGCTTTTTCAGG 1445
OY 168 HisIleGlyGlyLeuSerIleLeuAspProIlePheGlyValLeuSerAspValLeuThr 187
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1444 -----GCCCTCGTACAGCAGCAGCAGTGGCATTTTTCGGAATGAGCAGCAGC 1400
OY 188 AlaIlePheGlnAspThrValArg---LysGlnMetThrLysValLeuAlaProAlaPhe 206
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1399 CTAAATCGACCATGCAACATGGATATCGAGGAGAAAGAACCAATGCTGGCGGCGGAGG 1340
OY 207 LysArg 208
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1339 GCACGT 1334

```

RESULT 42

ABL15310/c
ID ABL15310 standard; cDNA: 4158 BP.

XX ABL15310;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 40412.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB71207.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

PS Claim 1; SEQ ID NO 40412; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB101840-AB16175) and the encoded proteins (AB157737-AB12072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4158 BP; 1104 A; 1022 C; 1036 G; 996 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Alignment Scores:

Pred. No.:	12.4	Length:	4158
Score:	85.00 <td>Matches:</td> <td>40</td>	Matches:	40
Percent Similarity:	38.17% <td>Conservative:</td> <td>31</td>	Conservative:	31
Best Local Similarity:	21.51% <td>Mismatches:</td> <td>67</td>	Mismatches:	67
Query Match:	7.96% <td>Indels:</td> <td>48</td>	Indels:	48
DB:	23	Gaps:	7

US-10-024-955-7 (1-213) x ABL15310 (1-4158)

```

OY 71 LeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAla 90
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2801 ATCGTAGTCCACACAGATGCTGTCGAGGCGTGCATGATCTGTCGCCAGGCTTGGC 2742
OY 91 AsnValLysGlyGluGlu-----GlyIleValAlaHisLeuLeuIle 105
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2741 CAGATCACCGTAGACACACGCTGCCAGAAAGTTGGCATCGAAGAGGACGCTTACGTA 2682
OY 106 GlyVal---HisAspAspIleValSerMetGlu-----TyrAspLeu 118
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2681 GGAGTAGCGGACAGAGATGATGTTGCCAGCGAGTAGAAGTAGCGGCGTTCATGTT 2622
OY 119 AlaTyrLysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPhe 138
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2621 GTGGGAAACCTGGCGCATGAT-----ATTGAAGCTGTGCAGATGTCCTCCCTCC 2568
OY 139 ValValAlaLeuSerLeuGluIleSerAspGluGlyAsnIleThrMetThrSerPheGlu 158
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2567 CTGGTCGAGCTGCTTCAATTCGCGACGAGCGGTAGTAGACAGAGCTGTTGAG 2508
OY 159 Val-----ArgGlnPheAlaAsnValValAsnHisIleGlyIleu----- 159
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2507 GTGATTTCTGTAGAGTCCCTGCCAGCTCCACCGCGGTGCTGCCAATAGATTG 2448
OY 160 -----ArgGlnPheAlaAsnValValAsnHisIleGlyIleu----- 172
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 2447 CTCCTCACATTGGCGGTGAGTTGCCACATTTGATGCGCATCTGTCACCTTTCCTG 2388

```

ABL15311/C
ID ABL15311 standard; cDNA; 2158 BP.
XX
AC ABL15311;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40415.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN M0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI: 2001-656860/75.
DR P-PSDB: ABB71208.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 40415; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2158 BP; 497 A; 622 C; 595 G; 444 T; 0 other;

Alignment Scores:
Pred. No.: 5.05 Length: 2158
Score: 85.00 Matches: 40
Percent Similarity: 38.17% Conservative: 31
Best Local Similarity: 21.51% Mismatches: 67
Query Match: 7.96% Indels: 48
DB: 23 Gaps: 7
US-10-024-955-7 (1-213) x ABL15311 (1-2158)
QY 71 leuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAla 90
DB 1801 ATCGATGCTCCACACGATCGGTGTCGACAGCGCATGATGTCGTGCCACAGGCTTGGC 1742
QY 91 AsnValLysGlyGluGlu-----GlyIleValLysAlaHisIleuLeu 105
DB 1741 CAGATCCACCGTAGACACCGCTGCCAGAGTAGTGGCATCGAAGAGGACGCTTACGTA 1682
QY 106 GlyVal---HisAspAspIleValSerMetGlu-----TyrAspLeu 118
DB 1681 GGAGTACGCGCAGAGATGATGTTCCCAAGGAGTAGAAGTAGCGCGTGCATCATGTT 1622
QY 119 AlaTyrLysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspHe 138

DB 1621 GTCCGGAAACTCGGGCGTATGCAT-----ATTGAAGCTGTACAGAGATGCCGCTCCTC 1568
QY 139 ValValAlaLeuSerLeuGluIleSerAspGlyAsnIleThrMetThrSerPheGlu 158
DB 1567 CTGGTCCAGCTGCTTCCAAATCGGCGACGCGTGTAGTAGACAGACGCTGTTCAG 1508
QY 159 Val-----SerIleLeuAspProIlePheGlyValLeuSerAsp----- 159
DB 1507 GTGATTTCTGTAGAAAGTCCCTGCCACGCTCCACCGCGTCCGTCCTCAATGAATTG 1448
QY 160 -----ArgGlnPheAlaAsnValValAsnHisIleGlyGlyLeu----- 172
DB 1447 CTCCGTCACATTTGGCGGTAGTCCGCCACATGTATGCGCATCTGTCACCTTGTCTG 1388
QY 173 -----SerIleLeuAspProIlePheGlyValLeuSerAsp----- 184
DB 1387 CAGCGATTTACACAGATCTCTAGATTGAGAGCGTTGTGCTGAGACCCGACGCTCCAG 1328
QY 185 ValLeuThrAlaIlePheGlnAspThrValArgLysGlnMetThrLysValLeuAlaPro 204
DB 1327 CTGACTCTGCAGCTGCTGGAAGACGCTGTGATGATACACAGCTCATCCGCGTGGCGCG 1268
QY 205 AlaPheLysArgGluLeu 210
DB 1267 GGAGTAGAAACTCTCTCTC 1250

RESULT 41
ABL21690/C
ID ABL21690 standard; DNA; 3277 BP.
XX
AC ABL21690;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16543.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN M0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 16543; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

PT producing animal models of audiogenic seizures -
 XX
 PS Claim 1: Page 38-45; 79pp; English.
 XX
 CC The present invention relates to nucleic acid molecules designated
 CC monogenic audiogenic seizure-susceptible-1 (mass1) genes. The nucleic
 CC acid molecule may be used via recombinant DNA methodologies in the
 CC production of transgenic animal (especially mouse) models for studying
 CC genetic abnormalities related to mass1 which result in seizure
 CC susceptible phenotypes (mass1 is audiogenic seizures). The present
 CC sequence is mouse monogenic audiogenic seizure-susceptible (mass1)
 CC cDNA.
 XX
 SQ Sequence 9437 BP; 2530 A; 1986 C; 2293 G; 2628 T; 0 other;
 Alignment Scores:
 Pred. No.: 33.2 Length: 9437
 Score: 85.50 Matches: 41
 Percent Similarity: 38.02% Conservative: 32
 Best Local Similarity: 21.35% Mismatches: 78
 Query Match: 8.01% Indels: 41
 DB: 22 Gaps: 8
 US-10-024-955-7 (1-213) x AAD18286 (1-9437)
 QY 26 ThrGluGluLeuAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThr 45
 DB 8317 ACCGTCAGACATGATCTGGCAACGACATGCGCAGAAATGTCAGCTTCAGACA 8376
 QY 46 IleAspPrometLysValProAspHisAlaAspLysPheGluArgHisVal----- 62
 DB 8377 GCTTCACATGTCATGATGGCCACAGACAGAAATGTCAGCTTCAGTGAAGACA 8436
 QY 63 -----GlyIleValAspPheLysGlyGluLeuAlaIleMetArgAsnIleGluAla 78
 DB 8437 CCCCAGGTCGAGGAATGTCACCTGCACTGCAACGTGGAAGTTGTTGACAAATGTAGAACGTC 8496
 QY 79 ArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlu----- 95
 DB 8497 -----AATTTGCTACTTACGCGGCCCACTCTTCTTC 8529
 QY 96 ---GluGlyIleValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMet 114
 DB 8530 TCTGAGGTACATGATGAATAAACATATTGTTGACATTGTCAGACAAATATCTCTGAG 8589
 QY 115 Glu-----TyrAspLeuAlaTyrLysLeuGlyAspLeuHis 126
 DB 8590 GAGAAAGAGTATACCAGGTTGTTCTGTATGATGTC-----AAGACCCCAAGAGTGTGCG 8643
 QY 127 ProThrThrHisValIleSerAspIleGlnAspPheValIleAlaLeuSerLeuGluIle 146
 DB 8644 CCAGCAGAGAGTGTCTACTTACTGATGCCAGGATATGCGAGCTGTACTGACATGAGACA 8703
 QY 147 SerAspGluGly-----AsnIleThrMetThrSer---PheGluValArgGln 161
 DB 8704 AGCGATGAGCCACACGAGTGTTTTAACATTGCTCTCTCCCTCAAGATTTGTTGCTCCAG 8763
 QY 162 PheAlaAsnValVal-----AsnHisIleGlyGlyLeuSerIleLeu 175
 DB 8764 GAGGCTAATGTAAACAATTACAGCTCTTCGTCACAGAGAGTTGCGATCTTAGACCAAC 8823
 QY 176 AspProIlePheGlyValLeuSerAspValLeuThr 187
 DB 8824 AATGTACAGTATGCTACTGTTCTCGAATAGTAAGT 8859
 RESULT 38
 AAV21210/c
 ID AAV21210 standard; DNA: 58407 BP.
 XX
 AC AAV21210;
 XX
 DT 10-NOV-1998 (first entry)
 XX

DE Methanococcus jannaschii large circular extrachromosomal element.
 XX
 XX Methanococcus jannaschii: methanogenic archaeon; circular chromosome;
 KW genome; autotrophic; extrachromosomal element; identification; ds.
 XX
 OS Methanococcus jannaschii.
 XX
 PN W09807830-A2.
 XX
 PD 26-FEB-1998.
 XX
 PF 22-AUG-1997; 97WO-US14900.
 XX
 PR 22-AUG-1996; 96US-0024428.
 XX
 PA (GENO-) INST GENOMIC RES.
 PA (UNIT) UNIV ILLINOIS FOUND.
 PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 PI Butl CJ, Smith HO, Venter JC, White OR, Woese CR;
 DR WPT, 1998-169145/15.
 XX
 PT Complete genome sequence of methano-genic archaeon, Methanococcus
 PT jannaschii - useful in identification of M. jannaschii genome
 PT fragment
 PS Claim 13; Page 585-600; 614pp; English.
 XX
 CC The present sequence represents the large circular extrachromosomal
 CC element sequence of the Methanococcus jannaschii circular chromosome.
 CC The present invention describes M. jannaschii open reading frames from
 CC the genome sequence. The invention also describes a computer based
 CC system for identifying fragments of the M. jannaschii genome that are
 CC homologous to target nucleotide sequences, comprising: (a) data storage
 CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550
 CC bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide
 CC sequence at least 99.9% identical to it; (b) search means for comparing a
 CC target sequence to the nucleotide sequence of the data storage means to
 CC identify a homologous sequence, and (c) retrieval means for obtaining
 CC the homologous sequence. The method, which is based on whole genome
 CC random sequencing of an autotrophic archaeon M. jannaschii, the genome
 CC of which consists of 3 physically distinct elements, a large circular
 CC chromosome (the 1664976 bp sequence given in AAV21209), a large circular
 CC extra-chromosomal element (the 58407 bp sequence given in AAV21210), and
 CC a small circular extra-chromosomal element (the 16550 bp sequence given
 CC in AAV21211), can be used in the identification of M. jannaschii genome
 CC fragment.
 XX
 SQ Sequence 58407 BP; 20163 A; 9034 C; 7447 G; 21761 T; 2 other;
 Alignment Scores:
 Pred. No.: 403 Length: 58407
 Score: 85.50 Matches: 51
 Percent Similarity: 39.53% Conservative: 34
 Best Local Similarity: 23.72% Mismatches: 82
 Query Match: 8.01% Indels: 48
 DB: 19 Gaps: 8
 US-10-024-955-7 (1-213) x AAV21210 (1-58407)
 QY 22 TyrAspLysIleThrGluGluLeuAsnLysAlaIleAspAspAlaIleAlaIleGlu 41
 DB 32250 TATTAAGAAATTTGTTAAAGAAATTTATGAAATGAGAAACAAAGAAATTTTAC 32191
 QY 42 GlnSerGluThrIleAspPrometLysValProAspHisAlaAspLysPheGluArgHis 61
 DB 32190 CAT-----ATTGATTTGAAATTTGTAAGGAAACATTAAGCAATCTCCACATTTG 32140
 QY 62 ValGlyIleValAspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeu 81
 DB 32139 GCAGAAATTAATGAA-----GAGCTAATTTCTTAATAAACATTTCTATGCCCTCATCA 32089

```

RESULT 36
AADI18294
ID AADI18294 standard; cDNA; 9436 BP.
XX
AC AADI18294;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mouse mass1 cDNA mutant (7009deltaG).
XX
KW Mouse; monogenic audiogenic seizure-susceptible-1 gene; mass1;
XX transgenic animal; genetic abnormality; seizure; mutant; ss.
XX
OS Mus musculus.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 796..7011
FT /tag= a
FT /product= "mouse mass1 protein mutant"
XX
PN M0200165927-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001MO-US06962.
XX
PR 03-MAR-2000; 2000US-187209P.
XX 03-AUG-2000; 2000US-222898P.
XX
PA (UTAH ) UNIV UTAH RES FOUND.
XX
PI Ptacek L, White S, Fu Y, Skradski S;
XX WPI; 2001-589903/66.
XX DR P-PSDB; AAE10949.
XX
PT Monogenic Audiogenic Seizure-Susceptible-1 (mass1) genes, useful for
XX producing animal models of audiogenic seizures -
XX
XX
XX Example 9; Page -: 79pp; English.
XX
XX The present invention relates to nucleic acid molecules designated
XX monogenic audiogenic seizure-susceptible-1 (mass1) genes. The nucleic
XX acid molecule may be used via recombinant DNA methodologies in the
XX production of transgenic animal (especially mouse) models for studying
XX genetic abnormalities related to mass1 which result in seizure
XX susceptible phenotypes (mass1 is audiogenic seizures). The present
XX sequence is mouse monogenic audiogenic seizure-susceptible (mass1)
XX cDNA mutant (7009deltaG). This mutant is constructed from the wild type
XX mass1 cDNA (SEQ ID NO:1) by deleting a nucleotide 'G' at position 7009
XX of the sequence.
XX
XX Note: The present sequence is not shown in the specification, but is
XX derived from the wild type mouse mass1 sequence (SEQ ID NO:1) shown in
XX the sequence listing of the specification.
XX
XX Sequence 9436 BP; 2530 A; 1986 C; 2292 G; 2628 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. NO.:
XX
XX Score: 33.2 Length: 9436
XX Percent Similarity: 85.50 Matches: 41
XX Best Local Similarity: 38.02% Conservatism: 32
XX Query Match: 21.35% Mismatches: 78
XX DB: 8.01% Indels: 41
XX Gaps: 8
XX
US-10-024-955-7 (1-213) x AADI18294 (1-9436)
OY 26 ThrcututlntleaknsyalsalaleaspalalelaalelaaleclugInSercIuThr 45
DB 8316 ACCGTGACAGCTCAACATCTTGCCAAACGACCAATGTGGCAGGAATTCGACGCTTCAGACA 8315

```

Oy	46	IIeAspPromeLysVal1ProAspRHisAlaAspLysPheLnuRghisVal-----	62
Db	8376	GCTTCAGATGTCATAGGCCACGAAGAAATGTTGCATCTCCATCTGGTAAGACA	8433
Oy	63	-----GlyIleValAspPheLysGlyLnuLeuAlaMetLArgAsnIleGluAla	78
Db	8436	CCCCAGGTCGAGAAATGTCACGTGCACCTGAAAGAGTTGGTTCGACAAATCTAGAAATGC	8495
Oy	79	ArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyLnu-----	95
Db	8496	-----AATTTCCTAACCTTTACGGGCCCACTCTTCTTC	8528
Oy	96	---GluGlyIleValLysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMet	114
Db	8529	TCTGAGGTCATCATGAAATAAACATATTTGTGCATTTGTGGATGACATATCTGTAG	8588
Oy	115	Glu-----TyrAspLeuAlaTyrLysLeuGlyAspLeuHis	126
Db	8589	GAGAAAGAGTATACCAAGTGTGTTCTGTATGATGTC-----AAGACCCAGAGAGTGTG	8644
Oy	127	ProthrThrHisValIleSerAspIleGlnAspPheValValAlaLeuSerLeuGluIle	146
Db	8643	CCACAGAGAGTTCCTCTACTGATGACCCAGAGATATGACAGTGTACTGACAGTGAAGCA	8700
Oy	147	SerAspGluGly-----AsnIlePheMetThrSer---PheLnuValArgin	161
Db	8703	AGCGATGAGCCACACAGGCTGTTTAACCTTGCTCTCCCTCCCAAGATTGTGTGCTCCAG	8760
Oy	162	PheAlaAsnValVal-----AsnHisIleGlyLysSerIleLeu	175
Db	8763	GAGGCTATATGACAAATTCAGCTCTTCGTCACACAGAGAGTTCGAGATCTAGAGCCATC	8822
Oy	176	AspProIlePheGlyValLeuSerAspValLeuThr	187
Db	8823	AATGTCACGATGCTACTGTTCTCTGCAATAGTAGT	8858
RESULT 37			
AD ID	AAD18286		
AC	AAD18286;		
XX	18-DEC-2001 (first entry)		
XX	Mouse monogenic audiogenic seizure-susceptible-1 (mass1) cDNA.		
XX	Mouse; monogenic audiogenic seizure-susceptible-1 gene; mass1;		
KW	transgenic animal; genetic abnormality; seizure; ss.		
OS	Mus musculus.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	796..9138	
FT		/*tag= a	
CT		/product= "Mouse mass1 protein"	
XX	W0200165927-A1.		
PN			
XX	13-SEP-2001.		
PD			
XX	02-MAR-2001; 2001WO-US06962.		
PF			
XX	03-MAR-2000; 2000US-187209P.		
PR	03-AUG-2000; 2000US-222898P.		
XX			
PA	(UTAH) UNIV UTAH RES FOUND.		
XX			
PI	Ptacek L, White S, Fu Y, Skradski S;		
XX	WPI; 2001-589903/66.		
DR	P-PSDB; AAE10924.		
XX			
PT	Monogenic Audiogenic Seizure-Susceptible-1 (mass1) genes, useful for		


```
Db 949 GTACGCGACCTGTGACCTGCGAGCGGAGCGGCGGCGACGACGTTCTTCCGCAC 1008
QY 97 ----- 97
Db 1009 CCAACCCATCGCTACCGCGGAGCGGCTTGACCTTGCGGGCTTCCCGGAGAAAGC 1068
QY 98 -----11eValylsAlaHisLeuIleGlyValHisAspAlaValSerMetGlu 115
Db 1069 GCGCGCACCGTGGGGGCGCATTCGCGGCGGTGATGATGATCTTGGGCGCGGAG 1128
QY 116 TyrAspLeuAlaTyrIleGlyValSerLeuHisProThrHisValIleSerAspIle 135
Db 1129 GCGGCGCTCCCGCATCGCGCGCGCATCTCGAACCGACCGCTCGGCTT----- 1176
QY 136 GluAspPheValAlaIleAlaSerLeuIleSerAspGluIleGlyValSerIleThr 155
Db 1177 CACGAGCAGCGCGGTAGCGCGGTGCTGAGCTT-----GATGCCGCTCTCGAC 1224
QY 156 SerPheGluValAlaArgIlePheAlaAsnValAlaAsnHisIleGlyValSerIleLeu 175
Db 1225 GAGCTGGCGCAAGCAGCTGTCGACGACGAACTTCCGACGTCGGGGCGAGCAGCAGCTC 1284
QY 176 Asp---ProIlePheGluValLeu 182
Db 1285 CAGTGTGATGTCCGAGCGCTGCTG 1308
```

RESULT 34

AAS54100/c
ID AAS54100 standard; DNA: 2832 BP.

AC AAS54100;

DT 13-FEB-2002 (first entry)

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #231.

KW Antisense; ds; prokaryotic cellular proliferation gene;

KM antibiotic; antibacterial; drug design.

OS Pseudomonas aeruginosa.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001MO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR WPI: 2001-611495/70.

DR P-PDB: AAU36241.

PT New polynucleotides for the identification and development of

PS antibiotics; comprise sequences of antisense nucleic acids -

PS Claim 27; Seq ID No 7737; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequences is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 2832 BP; 555 A; 942 C; 846 G; 489 T; 0 other;

Alignment Scores:

Pred. No.:	6, 38	Length:	2832
Score:	85.50	Matches:	55
Percent Similarity:	37.96%	Conservative:	27
Best Local Similarity:	25.46%	Mismatches:	106
Query Match:	8.01%	Indels:	28
DB:	23	Gaps:	8

US-10-024-955-7 (1-213) x AAS54100 (1-2832)

```
QY 6 LeuIleAlaIleAlaIlePheValAlaValSerAlaAspProIleHisTyrAspLysIle 25
Db 1404 CTCGTTGTGCGCGCGCGGCGAGCGTAGCAGACAGCTGCAT-----GACCACGCTC 1357
QY 26 ThrGluGluIleAsnLysAlaIleAspAlaIleAlaIleGluGlnSerGluThr 45
Db 1356 ACCGTTGACTGATGCGCGGTAGTGCAGCGGAGTGGTGCAGCAGCAGCGCTTGGC 1297
QY 46 IleAspProMetLysValProAspHisAlaAspLysPheGluArgHisValIleVal 65
Db 1296 ATGTCGCCGTTGACGTGCGGAAGATCGCGCGCTGATGATCTTGGCCACGTCGGT----- 1243
QY 66 AspPheLysGluIleAlaIleMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLys 85
Db 1242 -----GGCGTACTGCGTGCAGACCGCATCTCTGCGCGCTGGTGGTGAACCGAC 1192
QY 86 ArgGlnLysAspAlaAsnValLysGlyGluGluIleValIleValLysAlaHisLeu----- 103
Db 1191 CTGCTTTGTATCCACGATGATGATGTCACCGGTCGTGTAGCGCGGCTGTGCGCAT 1132
QY 104 LeuIleGlyValHisAspAlaIleValSerMetGluTyrAspLeuAlaTyrLysLeuGly 123
Db 1131 CTGGAAGCTTTCATGACACGCGCTGCCCGGAGCGGAGCATGCCGATGGAAT 1072
QY 124 AspLeuHisProThrHisValIleSerAspIleGlnAspPheValAlaIleLeuSer 143
Db 1071 CGTACCACCTTGTGCGCGGAGCTGTC-----CTTCCGCGGTCCTGCGGACGACGAC 1018
QY 144 LeuGluIleSerAspGluValSerIleThrMetThrSerPheGluVal----- 159
Db 1017 GGAACCTCGACACCGCGGGAAC-----GATTCACAGGTGGAGCGGTTGAA 970
QY 160 -----ArgGln-----PheAlaAsnValAlaAsnHisIleGly-----LeuSer 173
Db 969 GCGCAGCGCAGGTGAGTATCGCCACCGCTGTCATCATCATCGAGAGAAGACCGCTGTC 910
QY 174 IleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThr 193
Db 909 GTACTTCACGTCAACCGAGCCAGCATCTTCCCTTCGAACTGTCGAACAG 850
QY 194 Val-ArgLysGluMetThrLysValLeuAlaProAlaPheLysArg 208
Db 849 GTGCGCGGCTTGTGCGCAGGCTTGTGACGACGACGCTTACGCGA 804
```

RESULT 35

```

Db 727 GCGGTGCTGCTCCAGGTGAATCCGGACATCGCGCGGAGTGGCGGTAGCGGC 786
QY 34 AspaAlaIleAlaIleGlnGlnSerGluThrIleAspPro-----MetLys 50
Db 787 GGTCTGGAGATAGATGCGGCGGAGCAGCTCGAGGTGCGGATGATCGCGCGCGGAG 846
QY 51 ValProAspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGly--- 69
Db 847 GTGAGAGACCTGCGCGAT-----GCGCTTCTCGATCTTCTCGGCTC 888
QY 70 -----GluLeuAlaMetArgAsnIleGlnAlaArgGlyLeuLysGlnMet 84
Db 889 GATCTTGTGGGTGCGGATGCTCGACGAAGAGCGCGGCTCGGCTTGGCGATCGC 948
QY 85 LysArgGlnGlyAspAlaAsnValLysGlyGlnGluGly----- 97
Db 949 GTACGGGACCTGACCTCGACGCGAGCGGCGGCGGAGCAGCTTCTTCCGCAC 1008
QY 97 ----- 97
Db 1009 CAGCGCATGCGGTACGCGGCGGAGCGGCTCGACGCTTGGAGGCTTGGCGGAGAGGC 1068
QY 98 -----IleValLysAlaHisIleLeuIleGlyValHisAspAlaIleValSerMetGlu 115
Db 1069 GCGCGCACCGTGGCGGCGGCGCATGCCCGCGGTAGTGTGATGATCTTGGCGCGGTAG 1128
QY 116 TyrAspLeuAlaThrLysIleGluLysAspLeuHisProThrThrHisValIleSerAspIle 135
Db 1129 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
QY 136 GlnAspPheValValAlaLeuSerLeuGluIleSerAspGlnLysAsnIleThrMetThr 155
Db 1177 CAGGAGAGCGGGTAGCGGCTCGGTGTCGAGCTT-----GATGCGGCTGCTGAC 1224
QY 156 SerPheGluValArgIlePheAlaAsnValValAsnHisIleGlyLysSerIleLeu 175
Db 1225 GAGCTGGCAGACGATGCTCGACGAGACTTCGACGCTCGGCGGCGGAGCAGCGACTC 1284
QY 176 Asp---ProIlePheGlyValLeu 182
Db 1285 CAGCTGATGTCGAGGCGGTGTCG 1308

RESULT 33
AA256005
ID AA256005 standard; DNA; 1693 BP.
XX
AC AA256005;
XX
DT 23-MAR-2000 (first entry)
XX
DE Contig 004 from cosmid PKO5023-27 from Streptomyces venezuelae.
XX
KW Narbonolide polyketide synthase; PKS; cosmid PKO5023-27; contig 004;
KW ketolide; SAM synthase; S-adenosylmethionine synthase; hydroxylase;
KW picromycin; antibiotic production; narbomycin; ds.
XX
OS Streptomyces venezuelae.
XX
FH key Location/Qualifiers
FT CDS complement (694..1692)
FT /*tag= a
FT /*product= SAM synthase
FT /*note= "Partial S-adenosylmethionine synthase"
FT CDS complement (3..692)
FT /*tag= b
FT /*product= ORF16_product
FT /*note= "M. tuberculosis cbhK homologous partial protein"
XX
PN MO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.

```

```

XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
PI Ashley G, Belach MC, Belach M, McDaniel R, Tang L;
DR WPI; 2000-072618/06.
XX P-PDB; AAY67216, AAY67217.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
PS Disclosure; Page 38; 98pp; English.
XX
CC This is contig 004 from the recombinant cosmid PKO5023-27 DNA sequence
CC (see AA256001) which contains a Streptomyces venezuelae DNA insert. The
CC cosmid contains open reading frames which encode the various modules of
CC the narbonolide polyketide synthase (PKS). The invention relates to
CC recombinant DNA containing a coding sequence for a narbonolide PKS.
CC Polyketides are compounds synthesised from 2-carbon units through a
CC series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAT, PICATII, and PICATV). PICAI includes the loading module
CC and extender modules 1 and 2, PICAT includes extender modules 3 and 4,
CC PICATII includes extender module 5 and PICATV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid PKO5023-27. Narbonolide is desosaminylated in S.
CC venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC required for this conversion, and the desosaminyl biosynthetic genes are
CC also found in cosmid PKO5023-27. The recombinant DNA of the invention is
CC used to express, in transformed cells, narbonolide (or its derivatives)
CC or other ketolides (particularly hybrids), which may then be converted
CC (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC polyketide antibiotics or their intermediates. The antibiotics are useful
CC in human or veterinary medicine.
XX
SO Sequence 1693 BP; 237 A; 566 C; 633 G; 257 T; 0 other;

Alignment Scores:
Pred. No.: 3.16 Length: 1693
Score: 85.50 Matches: 46
Percent Similarity: 34.13% Conservative: 25
Best Local Similarity: 22.12% Mismatches: 84
Query Match: 8.01% Indels: 53
DB: 21 Gaps: 7

US-10-024-955-7 (1-213) x AA256005 (1-1693)
QY 14 AlaValSerAlaAspProIleHisTyrAspLysIleThrGlnIleAsnLysAlaIle 33
Db 727 GCGGTGCTGCTCCAGGTGAATCCGGACATCGCGCGGAGTGGCGGTAGCGGC 786
QY 34 AspaAlaIleAlaIleGlnGlnSerGluThrIleAspPro-----MetLys 50
Db 787 GGTCTGGAGATAGATGCGGCGGAGCAGCTCGAGGTGCGGATGATCGCGCGGAG 846
QY 51 ValProAspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGly--- 69
Db 847 GTGAGAGACCTGCGCGAT-----GCGCTTCTCGATCTTCTCGGCTC 888
QY 70 -----GluLeuAlaMetArgAsnIleGlnAlaArgGlyLeuLysGlnMet 84
Db 889 GATCTTGTGGGTGCGGATGCTCGACGAAGAGCGCGGCTCGGCTTGGCGATCGC 948
QY 85 LysArgGlnGlyAspAlaAsnValLysGlyGlnGluGly----- 97

```

The invention relates to evaluating strain variation within and between different populations of the tuberculosis bacterial pathogen, *Mycobacterium tuberculosis* or related *Mycobacterium* by determining the nucleotide sequence of the first strain at positions in the complete sequence of the genome that correspond to positions that differ in the nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AI199683) and H37Rv (AI199682). The method is useful for evaluating strain variation of *M. tuberculosis* and has valuable application in the fields of tuberculosis genetics, epidemiology, patient treatment and epidemic monitoring.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

QY 33 ILeaspharPhaIIealalalalelgluInsergluThrlIeasProMetLysValPro 52
:::||||| :::::
Db 1672916 TTGGACACCCCGCCGCGCGCGCGTGCAGACGCCCTGTATACAGGCCACCGATTGTGAAGGTGGCC 16728575GCC

QY 53 AsPhnIaIaAspLysPhnGluAArgHisValGlyIIeValaAspPhnLysGlyIuIuEnIIa 72
||| |||||::: ||| ||| |||
Db 1672856 GACCTGCAGCCCGGTGATATCGACCGCTATTCGGGGCCGCCGTGGCATGAAGCCGCAAGACC 16727979TTC

QY 73 MetArgAsnIIleGluIalAArgGlyLeuIysGlnMetLysArg----- 86
::: ::::: |||
Db 1672796 GCTCTGGACCTCGTCGACGGGGGGGCCACAGTCCCCCAAGGAACCTGGCTGGCTTGTG 16727373TTC

Qy	87	-----GInGlyspAlaAnVAlYgLyGluGluGly	97
Db	1672736	TTGATCGCGGCGCGGCTTTCGCGGCCCCAGACGCAGATCGCGGTGGCCAGATGGGGC	1672677
Qy	98	IlEAllyAlAnHsLsLeuIlEclYAlHIsApSAlIlEAlSMeTcIlGutyrAsp	117
		::: :::	:::
Db	1672676	TGGCGGAGACCCCTTTGGATATGGCGCTGGCAAGACATCAAGGTCGCGGCCGAGTATGAC	1672617
Qy	118	LeuAlATrLySLeuGlyAspLeuHIsPrOthrTrr-----HIsAlIlEsErSpIlE	135
		:::	
Db	1672616	-----GGCGACACACATCATCGACAGTGCCTACCACTACCGGAAAGATC	1672572

DT 22-JAN-2001 (first entry)
XX
XX Nucleotide sequence of ORF15 encoding S-adenosylmethionine synthetase.
DE
XX Narbonolide synthase; polyketide synthase gene; narbonolide polyketide
KM antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KM plicromycin biosynthesis; ss.

PN	US61.17659-A.
XX	

PD 12-SEP-2000.

PF 27-MAY-1999; 99US-0320878.
XY

PR	28-MAY-1998;	98US-0087080.
PR	22-SEP-1998;	98US-0100880.

PR	08-FEB-1999;	99US-0119139.
PR	20-MAY-1999;	99US-0134990.

PR	30-APR-1997;	97US-0846247.
PR	06-MAY-1998;	98US-0073538.

EN 20 AUG 1990, 0000 0141200.
XX
BT

XX	Bot'lach MC	Bot'lach M	Trans I	McDaniel B
PT	ach'low C			

XX WPI; 2000-610844/58.
DR

PT New recombinant pick hydroxylase gene of *Streptomyces venezuelae* useful

intermediates in the synthesis of compounds with pharmaceutical value

XX	
PS	Disclosure; Columns 45-46; 117pp; English

CC The present sequence is used to produce the recombinant DNA compounds

expressing recombinant polyketide synthase genes in host cells for the production of naphthalide derivatives and polyketides that

are useful as antibiotics and as intermediates in the synthesis of compounds with pharmaceutical value. The DNA compounds may also encode

cc a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
cc transferase enzymes (useful for conversion of ketolides to antibiotics)

CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity

CC host cells are useful as genetic systems that allow rapid engineering

creating novel ketolide analogs for pharmaceutical applications.

Sequence 1693 BP; 237 A; 566 C; 633 G; 257 T; 0 other;

Alignment Scores: 3 10 1000

Score:	85.50	Matches:	46
Percent similarity:	34.13%	Conservative:	25

Best Local Similarity:	22.12%	Mismatches:	84
Query Match:	8.01%	Indels:	53

DB:	21	Gaps:	7
-----	----	-------	---

US-10-024-955-7 (1-213) x AAA75637 (1-1693)

```

14 AlaValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIle 33
      |||||  ||  |||  :|||:  |||:
29

```



```

XX XX
PI PI  Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI PI  Yamamoto RT, Xu HH;
XX XX  WPI: 2001-611495/70.
DR DR  P-PSDB; AA036585.
XX XX
PT PT  New polynucleotides for the identification and development of
XX XX  antibiotics, comprise sequences of antisense nucleic acids -
PS PS  Claim 27: Seq ID No 8081; 511pp; English.
XX XX
CC CC  The invention relates to antisense inhibitors of genes essential to
CC CC  prokaryotic cellular proliferation, their use in identifying the
CC CC  genes' their use in the discovery of novel antibiotics, the essential
CC CC  genes themselves and the encoded proteins. The prokaryotes used are
CC CC  Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC CC  pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC CC  invention is also useful for the identification of potential new targets
CC CC  for antibiotic development. The antisense nucleic acids can also be used
CC CC  to identify proteins used in proliferation, to express these proteins,
CC CC  and to obtain antibodies capable of binding to the expressed proteins.
CC CC  The proteins can be used to screen compounds in rational drug discovery
CC CC  programmes. The antisense nucleic acid sequence is also useful to screen
CC CC  for homologous nucleic acids which are required for cell proliferation in
CC CC  a wide variety of organisms. The present sequence encodes an
CC CC  essential prokaryotic cellular proliferation protein.
CC CC  Note: The sequence data for this patent did not form part
CC CC  of the printed specification, but was obtained in electronic
CC CC  format directly from WIPO at
XX XX  ftp.wipo.int/pub/publ/published_pct_sequences.
XX XX
SQ SQ  Sequence 1995 BP: 770 A; 274 C; 394 G; 557 T; 0 other;

Alignment Scores:
Pred. No.: 3 Length: 1995
Score: 86.50 Matches: 36
Percent Similarity: 41.36 Conservative: 31
Best Local Similarity: 22.22% Mismatches: 50
Query Match: 8.10% Indels: 45
DB: Gaps: 5

US-10-024-955-7 (1-213) x AAS54444 (1-1995)
QY 1 MetMetlyshetleuilealaialaialaPhevalaIalValserAlaasproIle 20
DB 1063 ATCTTTTCGTTGTAATGATTCGCACTGTTCTTGTGGCAATGGCAATGTTGGTAAT 1122
QY 21 HistTyAspLysIleThrgIugIuIleasnLysAlaIleaspaspAlaIleAlaIle 40
DB 1123 AATACGAGAGACACCTGATGTAATCGGAAATCTGTAAAGACAGA----- 1170
QY 41 GluGlnSerGluThrIleaspPrometLysVal----- 51
DB 1171 GACCAATATTCATTAATAACAATCGAATAGTAATAATTTCTAGAACTTATAGTAT 1230
QY 52 -----ProaspIlaasLaspLysPheGluArg 60
DB 1231 AATATCCGAAAATGAAATTTAAGACACACCTCTAATACAGGTGACGTTGAACGT 1290
QY 61 -----HisValGlyIleValaaspPheLysGly----- 69
DB 1291 GCGACAGCTTGATGTTGTTATATCAAAAGCCCTGAAAGGTTAAATGCCAANTGTC 1350
QY 70 -----GluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMet 84
DB 1351 ATTTGTTTACCTTAAGGACCAAGCCTTCAGAAATTAATAATCGTTAGCTTAAAGATGTT 1410
QY 85 LysArgGlnGlyLysPalaAsnValLysGlyGlnGlyLe-----Val 99
DB 1411 AAGATTGAAAAGACTATATTAATTAATCAAGCCCAAAAGGATACATTGCAAAATGTTA 1470
QY 100 LysAlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyAspLeuAla 119

```

```

DB 1471 ACCGCAAAATCGAATCCGCTATTCAATGATTAATTAATTAATGATCTTACGC 1530
QY 120 TyrLys 121
DB 1531 ATTTAAG 1536
RESULT 29
AAH53069
ID AAH53069 standard; DNA; 934 BP.
XX
AC AAH53069;
XX
DR 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1531.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX
KW vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000MO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimerly WJ;
XX
DR WPI: 2001-316495/33.
DR P-PSDB; AAG82219.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 433; 2188pp; English.
XX
PS AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 934 BP: 322 A; 121 C; 166 G; 325 T; 0 other;

Alignment Scores:
Pred. No.: 1.22 Length: 934
Score: 86.00 Matches: 32
Percent Similarity: 46.09% Conservative: 27
Best Local Similarity: 25.00% Mismatches: 39
Query Match: 8.05% Indels: 30
DB: Gaps: 7

```



```

OY 49 MetLysValProAspHisAlaAspLysPheGluArgHisValGlyLeuValAspPheLys 68
   ::::::::::::::::::::
DB 124291 CTTAACTTCGAGGAAGCTGAGCTGATTAATAGAAAGACGGGGCTATATACAGATATGAAA 124350
OY 69 -----GlyGluLeuAlaMetArgAsnIleGluAla 78
   ::::::::::::::::::::
DB 124351 TTACGGAATTGTAAGCTACTCTCTCTGCGGGTTCCTCAACTCTCTCTCTCTCTCTCTCT 124410
OY 79 -----ArgGlyLeuLysGlnMetLysArgGlnGlyAspLysVal 92
   ::::::::::::::::::::
DB 124411 GTTCAATACCTCATATCCATCCATAGAAAGACGGGGCTCAACCATGGGAAGGCTTCTCTCA 124470
OY 93 LysGlyGluGluGlyLeuValLysAlaHisLeuLeuIleGlyValHisAspLysVal 112
   ::::::::::::::::::::
DB 124471 AAGGGGCTTGAGACAGCTTATCCAAATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 124506
OY 113 SerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThrHisValIle 132
   ::::::::::::::::::::
DB 124507 -----AACGTAATA 124515
OY 133 SerAspIleGlnAspPheValAlaLeuSerLeuGlnIleSerAspLysGlnIle 152
   ::::::::::::::::::::
DB 124516 TCAACGCTA-----ATA 124527
OY 153 ThrMetThrSerPheGluValArgGlnPhe-----AlaAsnValAlaAsnHisIleGlyGly 171
   ::::::::::::::::::::
DB 124528 GTGCGTAGTCGCAATTTCTATATCCGCAATTTCCGCCCATGATATATGAGCATTTGGCCGA 124587
OY 172 Leu-----SerIleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIle 189
   ::::::::::::::::::::
DB 124588 AACTAGAGAGCCAGACAGTCCCTCTTGGAAATCTCTGCTGATATATCTGGCCCTG 124647
OY 190 PheGlnAspThrValArgLysGlnMetThrLysValLeuAlaProAlaPhe 206
   ::::::::::::::::::::
DB 124648 ATTATAGGAGAGCTTCGCAATCGCAATGACCTCGTTATATGTTCCGCGCTAC 124698
RESULT 26
ABN68458
ID ABN68458 standard; DNA; 1308 BP.
XX
XX ABN68458;
XX
DE 01-JUL-2002 (first entry)
XX
DE Streptococcus polymucleotide SEQ ID NO 4829.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KM antiInflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus pyogenes.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tetfelin H;
XX
XX WPI: 2002-352536/38.
DR P-PSDB; ABP27827.
XX
PT New Streptococcus protein for the treatment or prevention of infection

```

```

PT or disease caused by Streptococcus bacteria, such as meningitis, and
PR for detecting a compound that binds to the protein -
PS Claim 7; Page 3645; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 1308 BP; 359 A; 258 C; 303 G; 388 T; 0 other:
XX
Alignment Scores:
Pred. No.: 1.47 Length: 1308
Score: 87.00 Matches: 47
Percent Similarity: 41.51% Conservative: 41
Best Local Similarity: 22.17% Mismatches: 80
Query Match: 8.15% Indels: 44
DB: 24 Gaps: 11
US-10-024-955-7 (1-213) x ABN68458 (1-1308)
OY 30 AsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIleAspPheMet 49
   ::::::::::::::::::::
DB 61 AATCGAATCGCAGAGAGCGCATTCATTTGAGAAAGCGTGAAGCGGCTTACTGCTGAC 120
OY 50 LysValProAspHisAlaAspLysPheGluArgHisValGlyLeuValAspPheLysGly 69
   ::::::::::::::::::::
DB 121 CGTATTTATGCTAGTGGGAGTGGCTTAACCGCAATTTCTCATTAATGATTAACAGCTGGT 180
OY 70 GluLeuAlaMetArgAsnIleGluAlaArgLysGlnMetLysArgGlnIle 88
   ::::::::::::::::::::
DB 181 -----ATTGATGACGCTGATGCCCGCTTATGAGCAATATCAGACAGCGCTCAA 231
OY 89 -----AspAlaAsn-----ValLysGlyGluGlyIleVal 99
   ::::::::::::::::::::
DB 232 ATTGCCATGGAAGAAGCTGATGTATGCTTTTGTGTTCTGTAAGAAAGGGGTGACC 291
OY 100 LysAlaHis-----LeuLeuIleGly 106
   ::::::::::::::::::::
DB 292 GACGCTGACGAATATGTGTAAAAATTCTGATTCGAACCAATATACCCAGCTCAATTTAGCT 351
OY 107 ValHisAsp---AspIleValSerMetGluTyrAspLeu-----AlaTyrLys 121
   ::::::::::::::::::::
DB 352 GTTAATAGCTTGAATATCCGAATGCGTATGACATTATGATTTTATCTACTGTGT 411
OY 122 LeuGlyAspLeuHisProThrThrHisVal-----IleSerAspIleGlnAsp 137
   ::::::::::::::::::::
DB 412 TTGGGTGATCCCTTATCCAGTTCTCTGTCATGCTATCGGAACAGGAGTCTCTTAGAT 471
OY 138 PheValValAla---LeuSerLeuGlnIleSerAspGlnGlyValSerIleThrMetThrSer 156
   ::::::::::::::::::::
DB 472 GCGATCGTTGAAGAACTGCGCACTAGAAAGCGTGAAGA---AATGATGATATTTTCCG 528
OY 157 PheGluValArgGlnPheAlaAsnVal-----ValAsnHisIleGly 171
   ::::::::::::::::::::
DB 529 TTCAGTTTAATCGCGTCTGCTAATGTGGGAATCTAGCTTATTAAGCGTATTTAGGT 588
OY 172 LeuSer-----IleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIle 189
   ::::::::::::::::::::

```


KW	kidney disease; diabetes; immunosuppressive disorder; otitis media;
KX	pneumococcal septicaemia; sinusitis; meningitis; therapy; ss.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	WO200006738-A2.
XX	
PD	10-FEB-2000.
XX	
PF	27-JUL-1999; 99WO-GB02452.
XX	
PR	27-JUL-1998; 98GB-0016336.
PR	19-MAR-1999; 99US-0125329.
XX	
PA	(MICR-) MICROBIAL TECHNICS LTD.
XX	
PI	Le Page RMF, Wells JM, Hanniffy SB, Hansbro PM;
XX	
DR	WPI: 2000-195301/17.
DR	P-PSDB; AAY81730.
XX	
PT	Streptococcal proteins and polynucleotides useful for diagnosis,
PT	treatment and prophylaxis of bacterial infections
XX	
PS	Claim 2; Page 48-49; 76pp; English.

CC This sequence encodes a *Streptococcus pneumoniae* protein of the
CC invention. The proteins (or their homologues, derivatives and/or
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
CC compositions comprising the proteins are useful as vaccines and also in
CC diagnostic assays. The sequences are useful for the detection or
CC diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested
CC with them. Agents capable of antagonising, inhibiting or interfering with
CC the function or expression of the protein or polypeptide are useful in
CC medical compositions in the treatment or prophylaxis of *S. pneumoniae*
CC infection. As the sequences can be used to treat *S. pneumoniae* infection,
CC they can be used to treat bacterial pneumonia, which has high rates in
CC young children, the elderly, and in patients with predisposing conditions
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC or with immunosuppressive disorders, especially AIDS. They can also be
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC meningitis.

SQ Sequence 1311 BP; 395 A; 269 C; 296 G; 351 T; 0 other;

Alignment Scores:	
Pred. No.:	1 12
Score:	88.00
Percent Similarity:	37.918
Best Local Similarity:	21.808
Query Match:	8.24%
DB:	21
Gaps:	8
Length:	1311
Matches:	46
Conservative:	34
Mismatches:	89
Indels:	42
Gaps:	8

US-10-024-955-7 (1-213) x AA291826 (1-1311)

Oy		30	Asysalalleasparallalelalleleluobinsgerlurhlleapromet	49
Db		61	AATGGATCGCTGGTGACCCGATTCATTGTAAABATGTGCAGGACGACCTGAC	120
Oy		50	LysValProaspHisAlaaspLysPhegluarghHisValcylIleValAspPheLysgly	69
Db		121	CGATTATTATGCCAACGGGTGAGTGGCTCAATCGTTCTTAGTCATGATTTGATACAGAGGA	180
Oy		70	GluLeuAlaMetArgasnIleGlualaargGlyLeuLysGlnMetLysArgGlnGlyAsp	89
Db		181	-ATTGATGATGTGTGATGCCTCCTTTCATGGAACAATCAACGACGACGACGAGA	231
Oy		90	Alaasn-----ValyselgylugluglyIleVal	99
Db		232	ATTGCCATGCAAGACGACATGTTATCGTTTTGTCTGTCTGTGTAAGGAAATAATTACT	291
Oy		100	Lys-----AlanHsleuLeuilecliyalaHisaspIleVal-----	112

```

Db 222 GATGACGACGAATACGTACGTCGTACGTTATATTAAGACCCACAACACGTTATCTCGCA 351
QY 113 -----Sermec1uYrAspleu-----AlatYrLys 121
Db 352 GTCAACAAGGTGGACAAACCCGTGACGAGAAATGATATATGATTTCTATGCTCGGT 411
QY 122 LeuGlyAspLeuHisProThrThHisVal-----11SerAsp11Leu1nAsp 137
      |||||::: ||| ::: ||| |||||::: |||
Db 412 TTGGGTGACACCATTCCTCATCTCATCTGTCATCGATGAATCGGTACAGGGATGTGTCGTGAT 471
QY 138 PheVal1Val1AlaLeuSerLeuGlu11SerAspGlu1nAsp11LeuThrMetThrSerPhe 157
      :::||||| :::||||| |||||::: |||
Db 472 GCGATCGTAGAAAAATCTTCCAAATGAATATGAGGAAGAAAATCCAGATGTGCATTAAGTTT 531
QY 158 GluVal1Arg1nGlnPheAlaAsnVal-----ValAsnHis11GluGlyLeu 172
      ::: ||||| |||||::: ||||| |||
Db 532 AGCTTGATTTGGTGTGCTCCCTACAAAGTTTGGAAGAAATCAAGCTTGATCATGCTATCTGGAGAA 591
QY 173 Ser-----11LeuAspPro11SerPheGlyVal1LeuSerAspVal1LeuThrAla11Phe 190
      ||| |||||::: ||| |||||::: |||
Db 592 GACCGTGTTATTTGCTAGACCTGCTGTGGAAACAACCTGTCATGCCATTTATACCCACTTT 651
QY 191 GlnAspThrVal1ArgLysGluMetThrLysVal 201
      ||||| :::|||| ||| :::
Db 652 ACAGATTACAGATGTGTCAAGACTTATACATGATTT 684

```

RESULT 24

AAS93828/c
ID AAS93828 standard; cDNA; 2700 BP.

AC AAS93828;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #29632

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food contaminants; medical; infectious diseases; genetic disorders;

XX
XX
HOMOCARD

XX
PN
W0200175067-A2

XX
PD 11-OCT-2001

XX
PF 30-MAR-2001: 2001WO-US08631

AA 31-MAR-2000; 2000US-0540217
PR

XX 23-AUG-2000, 2000005-004210/

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT,

DR WPI; 2001-639362/73

XX
XX
Nov 1 2014 10:11 AM

PT New Isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity -

PS Claim 1; SEQ ID NO 29632; 103pp; English

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as

```

OY 38 AlaAlaIleGluInSerGIuThrIle----- 46
DB 392 GGCACCATGAAGGTGGGGAGGTGGCCACATCACCTGCAACCAAGATATCCCTACGTT 441
OY 47 -----AspPrometLysValProAspHisAlaAsp---LysPheGluArgHisVal 62
DB 442 TCACGAGCGACGTCTCCCAAGATTCCTCCCAATGCCACGCTGTATTGAG-----GTG 495
OY 63 GlyIleValAspPheLysGlyIleuAlaMetArgAsnIleGluAlaArgIleuLys 82
DB 496 GAGTTGTTGAGTTTAAGGAGAGATCTGACGGAAGAGAATGGCGAATCATTCGCG 555
OY 83 GluMetLysArgGluGlyAspAlaAsnValLysGlyGluGlyIleValLysAlaHis 102
DB 556 AGAATACAGACTCGCGGTGAAGCATATGCTATAGCCCAATGACGCTATCGTGGAGGTT 615
OY 103 LeuLeuIleGlyValHisAspIleValSerMetGluTyrAspLeuAlaTyrLysIleu 122
DB 616 GCACCTGGAAGGTACTACAGAGACAAGCTCTTGAACGCGGAGCTCGCTTGGAGTT 675
OY 123 GlyAsp 124
DB 676 GCGGAG 681

RESULT 22
ID AAs54516 standard; DNA; 1311 BP.
XX AAs54516;
AC 11-APR-2001 (first entry)
XX
DE yphc gene of Streptococcus pneumoniae (GTP binding protein).
XX
KM yphc: GTP binding protein; antibody; treatment; infection;
KM bacteraemia; otitis media; conjunctivitis; pneumoniae;
KM meningitis; sinusitis; pleural empyema; endocarditis; ds.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 1..1311
FT /tag= a
FT /product= yphc protein (GTP binding protein).
XX
PN WO200068427-A1.
XX
PD 16-NOV-2000.
XX
PF 02-MAY-2000; 2000WO-US11894.
XX
PR 07-MAY-1999; 99US-0307003.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI Zalcain M, Biswas S, Burnham MKR, Warren PV, Ingraham KA;
PI Chalker AF, So CY, Holmes DJ, Warren RL, Van Horn S, Trainl CM;
DR WPI: 2000-687653/67.
DR P-PSDB; AAB04108.
XX
PT Streptococcus pneumoniae yphc protein and DNA sequence, useful for
PT treating infections, meningitis, and bacteraemia
XX
PS Claim 2; Page 37; 39pp; English.
XX
CC A new nucleotide is described which encodes a 436 residue
CC Streptococcus pneumoniae yphc (GTP binding protein). The DNA
CC sequence can be used to transform a host cell to produce
CC the protein and the products can be used to treat bacterial
CC infections (especially Streptococcus pneumoniae infections, and

```

```

CC Helicobacter pylori infections), otitis media, conjunctivitis,
CC pneumoniae, bacteraemia, meningitis, sinusitis, pleural empyema, and
CC endocarditis. The encoded protein and its peptide fragments can be
CC used to generate antibodies.
XX
SQ Sequence 1311 BP; 392 A; 273 C; 296 G; 350 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1-12 Length: 1311
Score: 88.00 Matches: 46
Percent Similarity: 37.91% Conservative: 34
Best Local Similarity: 21.80% Mismatches: 89
Query Match: 8.24% Indels: 42
DB: 21 Gaps: 8

US-10-024-955-7 (1-213) x AAs54516 (1-1311)
OY 30 AsnLysAlaIleAspAspAlaIleAlaIleGluInSerGIuThrIleAspPromet 49
DB 61 AATCGATGCGCTGTGAGGAGATCTCCATTGTAGAAGATGTGACAGATGACACGTGAC 120
OY 50 LysValProAspHisAlaAspLysPheGluArgHisValGlyIleValAspPheLysGly 69
DB 121 CGTATTATGCAACGCGGTGAGTGCCTCAATCGCTTTTACGATGATTGATACAGAGCA 180
OY 70 GluLeuAlaMetArgAsnIleGluAlaArgIleuLysGluMetLysArgGluGlyAsp 89
DB 181 -----ATTGATGATGTGATGCTCTCTTATGAGACAAATATACAGCAGCAGCA 231
OY 90 AlaAsn-----ValLysGlyGluGlyIleVal 99
DB 232 ATTCGCATGAGAGAGACAGATGTATCGTTTTTGTGCTGTGTAAGAGAGATTACT 291
OY 100 Lys-----AlaHisLeuIleGlyValHisAspIleVal----- 112
DB 292 GATGACAGAGATACGTAGCTGTGTAAGCTTTATAAGACCACAAACAGATTATCCTGCA 351
OY 113 -----SerMetGluTyrAspLeu-----AlaTyrLys 121
DB 352 GTCAACAAGGTGGACACACCTGATGATCCGAATGATATCTATGATTTCTATGCCCTCG 411
OY 122 LeuGlyAspLeuHisProThrThrHisVal-----IleSerAspIleGluAsp 137
DB 412 TTGGGTGAACACCTGCTATCTATCTGATCGTATGATGATGATGATGATGATGATGAT 471
OY 138 PheValValAlaLeuSerLeuGluIleSerAspGlyIleValIleThrMetThrSerPhe 157
DB 472 GCGATTGTGGAAACCTTCCAAATGATATGAAGAAAGAAATCCAGATGCTATTAAAGTTT 531
OY 158 GluValArgGlnPheAlaAsnVal-----ValAsnHisIleGlyGlyLeu 172
DB 532 AGCTTGATTTGCTGCTCCTACCTTGGAATATCACTGATTATGCTATCTTGGGAGCA 591
OY 173 Ser-----IleLeuAspProIlePheGlyValLeuSerAspValLeuThrAlaIlePhe 190
DB 592 GACCGTGCATTCGCGACGCTGTGCTGTAAGACACTCGTACGATGATGATGATGATGAT 651
OY 191 GluAspThrValArgLysGluMetThrLysVal 201
DB 652 ACAGATACAGAGTGAAGAGTTTACCATGATT 684

RESULT 23
ID AAs91826 standard; DNA; 1311 BP.
XX AAs91826;
AC 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae DNA sequence ID36.
XX
KM Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KM bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;

```

XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 2156 BP; 530 A; 551 C; 606 G; 469 T; 0 other;

Alignment Scores:
 Pred. No.: 1.92 Length: 2156
 Score: 88.50 Matches: 30
 Percent Similarity: 45.08% Conservative: 25
 Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.29% Indels: 17
 DB: 24 Gaps: 3

US-10-024-955-7 (1-213) x ABR84110 (1-2156)
 QY 18 ABRPRLIENHSTYRASPRLYSILETHRGUIGUILEASNLYSALIALEASPAALATIE 37
 DB 322 GACAAATTCCTCTTGACCTGGGAAAAGGAGGATCATCAAGGCTTGGACATTCGCATTA 381
 QY 38 AIAAIAIEGLUGLNSERGLUHRILE----- 46
 DB 382 GCCACCATGAAGTGGGGAGGTGCCACATCACCCTGCACAAACCAATATGCTACGCT 441
 QY 47 -----ASPRMELYSVALPROASPHISALASP---LYSPHEGLUHRHISVAL 62
 DB 442 TCAGCAGCGAGTCTCCAAAGATTCGCCCAATGCCACCTGTATTTAG-----GTG 495
 QY 63 GLYLIEVALASPHLYSGLYGULEUALAMETARGASNIIIGLUALARGLYLEULYS 82
 DB 496 GAGTGTGTTGATTTAAGGAGACAATCTGCACGGAAGGAAGATGGCGGAATCATTCGC 555
 QY 83 GIMELTYSARGLNGLYASPALAASVALYSGLYGULGUGLYLLEVALYLSALANIS 102
 DB 556 ACAATACAGACTCGCGGTACGCTATGCTAAGCCCAATGAGGCGTCTATCGTGGAGGCT 615
 QY 103 LEULLEIGLYVALHISASPRILEVALSERMETGLUTYRASPLEUALATYRYSLEU 122
 DB 616 GCACGTGAAGGGTACTACAAAGACAACATCTTTGACACGCGGAGCTCGCTTGGAGATT 675

QY 123 GLYASP 124
 DB 676 GCGCAG 681
 RESULT 21
 ID AAV24010
 AAV24010 standard; DNA; 2157 BP.
 AC AAV24010;
 XX 06-AUG-1998 (first entry)
 DT
 DE Human FK506 binding protein, FKBP52, coding sequence.
 XX
 KW FK506 binding protein; FKBP52; human; immune response regulator;
 immunosuppressant; steroid hormone receptor transformation; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 100..1479
 FT /*tag=a
 FT /product= FKBP52
 XX
 US5763590-A.
 XX
 PD 09-JUN-1998.
 XX
 PF 09-NOV-1994; 94US-0336618.
 XX
 PR 16-OCT-1992; 92US-0963325.
 PR 11-OCT-1991; 91US-0777752.
 PR 29-MAR-1994; 94US-0218989.
 PR 09-NOV-1994; 94US-0336618.
 XX
 PA (VERT-) VERTEX PHARM INC.
 XX
 PI Harding MW, Livingston DJ, Peattie DA;
 XX
 DR WPI; 1998-347419/30.
 DR P-PSDB; AAM54038.
 XX
 PT DNA sequence encoding human FK506-binding protein - and recombinant
 XX DNA molecule containing it
 PS
 PS Claim 1; Fig 3; 28pp; English.
 XX
 CC This sequence encodes 52 kD human FK506 binding protein, referred to as
 CC FKBP52, of the invention. The FKBP52 protein plays a key role in
 CC regulating immune responses. FKBP52 may be useful for mediating steroid
 CC hormone receptor transformation. The DNA may be used to screen for new
 CC immunosuppressants, and in assays for metabolites in samples from
 CC individuals taking immunosuppressants. The DNA may also be used in
 CC assays for identifying natural intracellular rapamycin-like or FK506
 CC like substances, and in assays for identifying natural intracellular
 CC substrates that are potential targets for other immunosuppressants.
 XX
 SQ Sequence 2157 BP; 531 A; 551 C; 606 G; 469 T; 0 other;

Alignment Scores:
 Pred. No.: 1.92 Length: 2157
 Score: 88.50 Matches: 30
 Percent Similarity: 45.08% Conservative: 25
 Best Local Similarity: 24.59% Mismatches: 50
 Query Match: 8.29% Indels: 17
 DB: 19 Gaps: 3

US-10-024-955-7 (1-213) x AAV24010 (1-2157)
 QY 18 ABRPRLIENHSTYRASPRLYSILETHRGUIGUILEASNLYSALIALEASPAALATIE 37
 DB 322 GACAAATTCCTCTTGACCTGGGAAAAGGAGGATCATCAAGGCTTGGACATTCGCATTA 381

Db	273	-----TTCTTATGTCGTCGCGACCTTGGACGACGAGCACC	226
RESULT	19		
ID	AAF21640		
XX	AAF21640 standard; DNA; 1903 BP.		
XX			
AC	AAF21640;		
XX			
DT	27-MAR-2001 (first entry)		
XX			
DE	Human breast and ovarian cancer associated antigen gene SEQ ID 27.		
KW	Human; breast cancer; ovarian cancer; cytosolic; immunosuppressive; neurotrophic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; immune disorder; Addison's disease; allergy; autoimmune hemolytic anaemia; autoimmune thyroiditis; diabetes mellitus; Crohn's disease; multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disorder; wound healing; neurological disease; ds.		
KM			
XX			
OS	Homo sapiens.		
XX			
PN	WO20005173-A1.		
XX			
PD	21-SEP-2000.		
XX			
XX	08-MAR-2000; 2000WO-US05881.		
PF			
PR	12-MAR-1999; 99US-0124270.		
XX			
FA	(HUMA-) HUMAN GENOME SCT INC.		
XX			
PI	Rosen CA, Ruben SM;		
XX			
DR	WPI: 2000-611515/58.		
DR	P-PSDB; AAB58737.		
XX			
PT	New human breast and ovarian cancer associated gene sequences and the polypeptides encoded by these genes, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases -		
PT			
PS	Claim 1; Page 501; 1299pp; English.		
XX			
CC	Sequences AAF21614 - AAF22031 represent DNA sequences encoding human proteins AAB58711 - AAB59128. The DNA and protein sequences are associated with breast and ovarian cancer. Included in the invention are sequences AAF22032 - AAF22040 and AAB59129 which are used in the isolation and characterisation of the DNA and protein sequences of the invention. The breast and ovarian cancer associated DNA, protein, agonist or antagonist sequences exhibit cytostatic; immunosuppressive; neurotrophic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic and cardiant activity. The polynucleotide and protein sequences are used in the diagnosis of cancer, particularly breast and ovarian cancer. The nucleic acid sequences, proteins, agonists and agonists may also be used in the diagnosis, prevention and treatment of immune disorders e.g. Addison's disease, allergies, autoimmune hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular disorders such as myocardial ischaemias; wound healing; neurological diseases such as cerebral anoxia and epilepsy; and infectious diseases.		
CC			
XX			
SD	Sequence 1903 BP; 478 A; 470 C; 528 G; 427 T; 0 other;		
Alignment Scores:			
Pred. No.:	1.62	Length:	1903
Score:	88.50	Matches:	30
Percent Similarity:	45.08%	Conservative:	25
Best Local Similarity:	24.59%	Mismatches:	17
Query Match:	8.29%	Indels:	17

DB:	21	Gaps:	3
US-10-024-955-7 (1-213) x AAF21640 (1-1903)			
OY	18 AspproIleHISlyrAspLysIleThrGluGluIleAsnLysAlaIleAspAlaIle 37		
Db	62 GACAAATTCCTCTTTGACCTTGAGAAAGGAGGTGATCAAGGCTTGAGACATTGCCATA 121		
OY	38 AlaIaIleGluInSerGluThrIle----- 46		
Db	122 GCCACCATGAAGGTGGGGAGGTGGCCACATCATCTGCACAACGAAATATGCGTACGGT 181		
OY	47 -----AsppromEtlYsValProAspHisAlaSp---LysPheGluAtrGHisVal 62		
Db	182 TCAGCAGGACCTCTCCCAAAATTCCTCCCAATCCCGCCATCCAGCGCTTGATTTAG-----GTG 235		
OY	63 GlyIleValAspPheLysGlyIleAlaMetLArgsnIleGluAlaIArgIleLys 82		
Db	236 GAGTTGTTGGATTAAAGGGAAGACTCTGACGGAAGAGAAAGTGGCGAATCATTCGCG 295		
OY	83 GlnMetLysArGInGlyAspAlaAsnValLysGlyGluGluGlyIleValAlaHis 102		
Db	296 AGAATACGAGCTCGCGGTGAAGGCTATGCTTAAGGCCCAATGAGGCTGTATCGTGAAGTT 355		
OY	103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeu 122		
Db	356 GCATCGGAAGGGTACTACAAAGCAAGCTCTTGACCACGCGGGAGCTCCGCTTGAGATT 415		
OY	123 GlyAsp 124		
Db	416 GCGGAG 421		
RESULT 20			
ID	ABK84110		
XX	ABK84110 standard; cDNA; 2156 BP.		
XX	ABK84110;		
DT	14-AUG-2002 (first entry)		
DE	Human cDNA differentially expressed in granulocytic cells #681.		
XX	Human; ss; granulocytic cell; DNA chip; bacterial infection;		
KW	viral infection; parasitic infection; protozoal infection;		
KW	fungal infection; sterile inflammatory disease; psoriasis;		
KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;		
KW	cardiac reperfusion injury; renal reperfusion injury; ARDS;		
KW	adult respiratory distress syndrome; inflammatory bowel disease;		
KW	Crohn's disease; ulcerative colitis; periodontal disease;		
XX	granulocyte activation; chronic inflammation; allergy.		
OS	Homo sapiens.		
XX			
PM	WO200228999-A2.		
XX			
PD	11-APR-2002.		
XX			
PF	03-OCT-2001; 2001WO-US30821.		
XX			
PR	03-OCT-2000; 2000US-237189P.		
XX			
PA	(GENE-) GENE LOGIC INC.		
PI	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;		
DR	WPI; 2002-435328/46.		
XX			
PT	Detecting granulocyte activation by detecting differential expression		
PT	of genes associated with granulocyte activation, which serves as		
PT	diagnostic markers that is useful for monitoring disease states and		
PT	drug toxicity -		
PS	Claim 1; SEQ ID NO 681; 114BP; English.		

```
OY 7 IleaIaIaValaIaPheValaIaValaSerAlaAspProIleHisTyrAspIleThr 26
    ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1804 ATTATGGCCCTACCAACATATTCGATTGTAGACGTC--AAATGGTGAATCAAC 1748
OY 27 GIUGIUIleAsnIleAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIle 46
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1747 ---CTATTAACTGATCGCTGTGAGCAATCTCCATTGTAGAAAGTCCGACAGAG 1691
OY 47 AspProMetIleValProAspHisAlaAspIlePheGluArgHisValGlyIleValAsp 66
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 1690 ACACGTGACCGGATTATTATGCAACGGGTGAGTGCCTCAATCGTCTTTTATGCAATGATGAT 1631
OY 67 PheIysGlyGluIleuAlaMetArgAsnIleGluAlaArgGlyIleuIysGlnMetIysArg 86
    ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1630 ACAGAGAGA-----ATGATGATGTGCAATGCTCTTATGTAAGCAAAATCAAGCAC 1580
OY 87 GInGlyAspAlaAsn-----ValIysGlyGluGlu 96
    ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1579 CAGGCAGAAATGGCCATGGAGAGACAGATGTTATCGTTTGTGCTGTCTGTAAAGAA 1520
OY 97 GIYIleValIys-----AlaHisIleuIleGlyValHisAspAspIle 111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1519 GGAATTACTGATGCAGACGAATACGTAGCTTAAGCTTTATTAAGACCCACAACCAAGTTC 1460
OY 112 Val-----SerMetGluTyrAspIle----- 118
    ::: ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1459 ATCCGCGAGTCACAGAGGTGACAAACCTGAGATGAGAAATGATATATGATTTCAT 1400
OY 119 AlaTyrIysIleuGlyAspIleuHisProThrThrHisVal-----IleSerAsp 134
    ||| ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1399 GCTTCGCGTGGTGGTGAACCATTTGCTCATCTGTCTCATGATCGATCGATCGATCGATCG 1340
OY 135 IleGlnAspPheValValAlaIleuSerIleuGluIleSerAspGluGlyAsnIleThrMet 154
    ::: ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1339 GTCGTAGATGCCATGCTAATAAATCTTCCCAATGAATATGAGAGAAATCCAGATGTC 1280
OY 155 ThrSerPheGluValArgIlePheAlaAsnVal-----ValAsnHisIle 169
    ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 1279 ATTAAATTAGCTGATGATGCTGCTGCTCAACGTTGGAAATCAAGTGTATCATGCTATC 1220
OY 170 GlyGlyIleuSer-----IleIleuAspProIlePheGlyValIleuSerAspValIleThr 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1219 TTGGAGAAAGACCGTGTATTGCTAGTCTGTGCGAGAACCACTCGTATGCCATTGAT 1160
OY 188 AlaIlePheGlnAspThrValArgIysGlnMetThrIysVal 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1159 ACCCACTTACAGATACAGATGTGTCAAGATTACCATGATT 1118

RESULT 18
AA048032/c
ID AA048032 standard; DNA; 1131 BP.
XX
AC AA048032;
XX
DT 08-FEB-1994 (first entry)
XX
DE FInd DNA.
XX
KW N-terminal; minor; fimbrial subunit; B. pertussis; fimbriae; FInd;
  vaccine; whooping cough; immune response; B. parapertussis;
  B. bronchiseptica; ss.
XX
OS Bordetella pertussis.
XX
FH Key Location/Qualifiers
FT sig_peptide 1..111
FT /*tag- a
FT mat_peptide 112..1131
FT /*tag- b
XX
XX EP555894-A.
XX
```

```
PD 18-AUG-1993.
XX
XX 08-JAN-1993; 93EP-0200047.
XX
PR 08-JAN-1992; 92EP-0200038.
XX
PA (NEW-) NEDERLANDEN MIN WELZIJN.
XX
PI Mool FR.
XX
DR WPI: 1993-260156/33.
DR P-PSDB; AAR40270.
XX
XX Vaccines against whooping cough - contains functional component
PT of fimbriae of B. pertussis and can also be used in veterinary
PT applications
XX
XX Claim 2; Page 12-15; 25pp; English.
XX
XX This sequence encodes the minor fimbrial subunit of B. pertussis
CC fimbriae (FInd). FInd and its corresponding DNA may be used to
CC produce a vaccine against whooping cough. It may induce an effective
CC immune response against all B. pertussis strains esp. B. parapertussis,
CC the second most important causative agent of whooping cough. The
CC vaccine may also be used to protect against B. bronchiseptica
CC infections.
XX
SQ Sequence 1131 BP; 212 A; 365 C; 359 G; 195 T; 0 other;

Alignment Scores:
Pred. No.: 0.795 Length: 1131
Score: 88.50 Matches: 33
Percent Similarity: 36.67% Conservative: 33
Best Local Similarity: 18.33% Mismatches: 61
Query Match: 8.29% Indels: 53
DB: 14 Gaps: 6

US-10-024-955-7 (1-213) x AA048032 (1-1131)
OY 14 AlaValSerAlaAspProIleHisTyrAspIleThrGluGluIleAsnIysAlaIle 33
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 GCGCTCGCGCTGCGTCCCGCCGATGA-----AACCACTTGTA 616
OY 34 AspAspAlaIleAlaIleAlaIleGluIleuSerGluThrIleAspProMetIysValProAsp 53
    ::: ::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 615 GGTGAACAGCTTGGAACTGGCACAGCACAGCTTGGCGCGCGCTGAGTGGCCGAC 556
OY 54 HisAlaAspIysPheGluArgHisValGlyIleValAspPheIysGlyGluIleuAlaMet 73
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 555 ---CTTGATCGCCTGATACGAACCTCCGCTTGTGCGGCTGATGATGAGCAGCTCGAA 499
OY 74 ArgAsnIleGluAlaArgIleuIysGlnMet-----LysArg 86
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 498 AAGCACTTGGCTTCTGTGGGATACAGCCACTGCCGCTTCCGACCTTGCATCGGCA 439
OY 87 GInGlyAspAlaAsnValIysGlyGluGluGlyIleValIys----- 100
    ::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 438 ACTGCCGAATGCGATATATAGATGCCGATGCCGCTCTCCGAGGAGCCAGCGCAATGC 379
OY 101 -----AlaHisIleuIleGlyValHisAspAspIleValSerMetGluTyrAsp 117
    ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
DB 378 GACCAAGATGCCCGCATGACATAGACACGCGCGGATCTTGTA----- 334
OY 118 LeuAlaTyrIysIleuGlyAspIleuHisProThrThrHisValIleSerAspIleGlnAsp 137
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 333 ----- 301
OY 138 PheValValAlaIleuSerIleuIleSerAspGluGlyAsnIleThrMetThrSerPhe 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 300 TTGCGCGTCAATGCTGTGCTGAGGTC----- 274
OY 158 GluValArgGlnPheAlaAsnValValAsnHisIleGlyIleuSerIleIleuAspPro 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

US-10-024-955-7 (1-213) x AAS89799 (1-3496)

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	3.25	3496	51	41	80	62
Percent Similarity:	89.00					
Best Local Similarity:	39.32%					
Query Match:	21.79%					
DB:	23					10

US-10-024-955-7 (1-213) x AAS89799 (1-3496)

QY 6 LeuIleAlaValAlaPheValAlaValSerAlaAsp----- 18

DB 650 GTGTGGGAGTCCCGCGCTGATTAACAGTACGTATGCGCTGAATTTGTTG 709

QY 19 -----ProIleHis-----TyrAspLysIleThrGluIleAsn----- 30

DB 710 CCGGGGATGTTCCCGACCATATTCGTTACAGCGCGCTGCTGAATGAAGAACAATT 769

QY 31 -----LysAlaIleAspAlaIleAlaIleGluIleSer 43

DB 770 ACCAAACCTTCGCGAGTGTGAGGCGATGATTAACGTCTGCTGCGTTGAATCTGGC 829

QY 44 GluThrIleAsp-----PrometLysValPro 52

DB 830 GAATCGTCTCACTTTGGGGGAAATGGCTGTGAATCAACCTGTGAAGTGTG 889

QY 53 AsnHisAlaAspLysPheGluValGlyIleValAspPheLysGlyIleValAla 72

DB 890 TGTGTATTATCCCATGGCTCCACGAGCGAAATATTATTTTCGGGAGAA----- 943

QY 73 MetArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnVal 92

DB 944 -----GAGATTCAAGGAGTCACATCCGCGATACCGAAGCAAGATATCGCATCAT 997

QY 93 LysGlyGluGluGlyIleValLys-----AlaHisLeuLeuIleGly--- 106

DB 998 CATCAAGATTTGGCCCTGTGAAGAAATGACCGTGTGAAGAAATTTCTCGGTAC 1057

QY 107 -----ValHisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeuLys 124

DB 1058 GAATTAACCCACAAATGGCAT-----ATGCAATTATGACCTG----- 1093

QY 125 LeuHisProThrThrHisValIleSerAspIleGlnAspPheValAlaLeuSerLeu 144

DB 1094 -----ATGACGCTACGCTGTCAAGACGTCGACAGAGTCAAGTTTA 1135

QY 145 GluIleSerAspGlu-----GlyAsnIleThrMetThrSerPheGluValArgGln 161

DB 1136 TCCATTTCACGTATACCCCGCGTGTGGCATTTTGGGCAACAACTGGTTGAA 1195

QY 162 PheAlaAsnValAlaValAsnHisIleGlyLeuSerIleLeuAspProIlePheGlyVal 181

DB 1196 ATGGCAAGGACCTTAATTAACAGGTGCGCTTGAATTTTCGATGAACCGACGCTCA 1255

QY 182 LeuSerAspValLeuThrAlaIlePheGlnAspThrValArg 195

DB 1256 TTACTGAGCAGAAACGTGATTTTACTGATATTATTCGC 1297

RESULT 17

AAV52212/C

ID AAV52212 standard: DNA: 5066 BP.

AAV52212;

23-OCT-1998 (first entry)

Streptococcus pneumoniae genome fragment SEQ ID NO:79.

Streptococcus pneumoniae; S. pneumoniae; genome: diagnosis; assay;

computer readable medium; vaccine; pharmaceutical composition; ds.

Streptococcus pneumoniae.

W09818931-A2.

07-MAY-1998.

30-OCT-1997; 97WO-US19588.

31-OCT-1996; 96US-0029960.

(HUMA-) HUMAN GENOME SCI INC.

Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

Kunsch CA, Rosen CA;

WPI; 1998-272225/24.

Claim 1; Page 634-636; 1409pp; English.

The present invention describes a computer readable medium which has

the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)

recorded on it, or a representative fragment or a sequence at least 95%

identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in

SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from

Streptococcus pneumoniae. The present invention also describes an

isolated nucleic acid molecule encoding a homologue of any of the

fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the

nucleic acid molecule is produced by a process comprising: (a) screening

a genomic DNA library using as a probe a target sequence defined by any

of the sequences in SEQ ID NO:1 to 391, identifying members of the

library which contain sequences that hybridise to the target sequence and

isolating the nucleic acid molecules from the members; or (b) isolating

mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid

molecules whose nucleotide sequence is homologous to amplification

primers derived from the fragment of the S. pneumoniae genome to prime

the amplification and isolating the amplified sequences. The computer

readable medium can be used in a computer-based system for identifying

fragments of the S. pneumoniae genome of commercial importance, or

expression modulating fragments of the S. pneumoniae genome. Products

from the present invention can be used in diagnosis kits and assays, and

pharmaceutical compositions and vaccines for S. pneumoniae.

Sequence 5066 BP; 1395 A; 1145 C; 983 G; 1543 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	5.4	5066	52	38	100	44
Percent Similarity:	89.00					
Best Local Similarity:	38.46%					
Query Match:	22.22%					
DB:	19					10

US-10-024-955-7 (1-213) x AAV52212 (1-5066)

ID	ABK50883	standard; DNA; 5529 BP.
XX	ABK50883:	
DT	30-JUL-2002	(first entry)
DE	DNA encoding Plasmodium falciparum liver stage antigen-3 (LSA-3) #1.	
KW	Liver stage antigen-3; LSA-3; vaccine; Th1-inducing adjuvant;	
KM	malaria parasite; malaria; protein-specific cytotoxic T cell response;	
XX	gamma-interferon; gene; ds.	
OS	Plasmodium falciparum.	
FH	Key	Location/Qualifiers
FT	CDS	1..5529
FT		/tag= a
FT		/product= "LSA-3"
FT		/note= "Liver stage antigen-3"
FT	exon	/transl_except= (pos:5524..5526, aa:Lys-Glx)
FT		1..198
FT		/tag= b
FT	Intron	/number= 1
FT		199..366
FT		/tag= c
FT	exon	/number= 1
FT		367..5529
FT		/tag= d
FT		/number= 2
PN	EPI201250-A1.	
XX		
PD	02-MAY-2002.	
XX		
PF	25-OCT-2000; 2000EP-0203724.	
PR	25-OCT-2000; 2000EP-0203724.	
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.	
PA	(INSP) INST PASTEUR.	
PI	Cohen J, Druilhe P;	
DR	WPI: 2002-373883/41.	
DR	P-PSDB; AAU96699.	
PT	Vaccine for treatment or prevention of malaria, comprises liver stage	
PT	antigen and adjuvant that induces Th1 response -	
PS	Disclosure; Page 16; 56pp; English.	
CC	The invention describes a vaccine comprising a Th1-inducing adjuvant (I)	
CC	and a protective liver-stage antigen (LSA), or its immunogenic fragment,	
CC	from a human malaria parasite, where LSA is a fragment of LSA-3, then (I)	
CC	is not montanide. The vaccine is useful for treating or preventing	
CC	malaria, specifically where caused by Plasmodium falciparum. The vaccine	
CC	is formulated in SBAS2 adjuvant, an oil-in-water emulsion containing QS21	
CC	and 3D-WPL (de-O-acetylated monophosphoryl-lipid A, the combination of	
CC	which results in strong induction of a circumsporozoite protein-specific	
CC	cytotoxic T cell response, which is not generally induced by vaccines	
CC	based on recombinant proteins, and synergistically increases production	
CC	of gamma-interferon. This sequence encodes the liver-stage antigen-3	
CC	(LAS-3) from the Plasmodium falciparum strain KI used to develop the	
CC	recombinant proteins for the vaccine.	
CC	Sequence 5529 BP; 2473 A; 435 C; 1174 G; 1447 T; 0 other:	

Alignment Scores: 4.62 Length: 5529

Pred. No.: 90.00 Matches: 47

Percent Similarity:	44.55%	Conservative:	43
Best Local Similarity:	23.27%	Mismatches:	78
Query Match:	8.43%	Indels:	34
DB:	24	Gaps:	9

US-10-024-955-7 (1-213) x ABK50883 (1-5529)

QY	22	TYTASRPLYSIIETHGUGLUILLASNLYSALALEASPRALAIIEIAIAIEIGLU	41
DB	3103	TTTTAATCTGATTTAGATAAAGTAGAGAAACACATAGAAATTTACGGAGAAAGTTTAGAA	3162
QY	42	GLNSERGIVTPIIIEASPROMETEUSVALPROASPHISALIASRPLSPHEGLUARGHIS	61
DB	3163	AACCATGA-----ATGCAATTAACCATTTTTTTTACTGCAAAATTTTGATAATGTA	3210
QY	62	VALGIIILEVALASRPHLEYSGLYGLUENALAMELARGASNIIEGLUALAARGLYLEU	81
DB	3211	AAAGCAATACAGAAATAATTTATTAACAGGTGATTTTCGAACATATAGAAACACATAGTA	3270
QY	82	LYSGIMELYSATRGINGLYASPRALASNAVILYSGLYGUGLYILEVALYALA	101
DB	3271	ATCCCATACAGAAAGAAAGTTGATTTGAAT-----GAAATATGGGTAGTTGCG	3318
QY	102	HISLEULEUILEGLYVALHISASPRAPRIEVALSERMETGIVTYSRPLEUALATYFLYS	121
DB	3319	ATTTTA-----GATATATATAGAAAATATGAAAGAGGTTTATTTAAATAAA	3363
QY	122	LEUGLASRPLEUHIISPROTHRTTHNISVALIIESEASPIIEGIIASRPHIVALAIA	141
DB	3364	TTAGAAATATTTTCACTACACTGAAAGGTGTCAGAAACSTGTAACATGTA-----	3417
QY	142	LEUSERLEUGLIUIIESERASPLUGLYASNIIEHMETHRSERPLEGVAL-----	159
DB	3418	-----GAACAAATAGTA--TATGTGATGTTGATGTTCTGCT	3453
QY	160	-----ARGINPHEALASNAVIALASNHISILEGLYGLYLEUSER-----	173
DB	3454	ATGAAGATCAATTTTATAGCAATATTAATATAGCAGAGGCGTTGAAAGAAATGTTTTT	3513
QY	174	IIELEUASPROIIEPHEGLYVALLEUSERASPVALLIETHRALA-----IIEPHEGLN	191
DB	3514	AATTTGGAAGATGATTTTAAAGTCGAACGATGTAATTTACTGTAGAAGAAATTAAGAT	3573
QY	192	ASPHITVALAIGLYSGIMETHRLYVALLEUALAPROALAPHELYSATRGIVLEUGLU	211
DB	3574	GAACCGGTTCAAAAAGAGTAGAAAAGAAACSTGTAGATTATTAGAA---GAATGGA	3630
QY	212	LYSASN	213
DB	3631	GAATA	3636

RESULT 15

AC	AAAT78867	standard; DNA; 6152 BP.
XX	AAAT78867;	
DT	08-OCT-1997	(first entry)
XX		
DE	P. falciparum	liver stage antigen-3 genomic sequence.
XX		
KW	Plasmodium falciparum;	pre-erythrocyte; liver stage antigen; serum;
KV	prophylaxis; Thai strain;	gene organisation; exon; intron; hydrophobic;
KW	glycosyl-phosphatidylinositol	membrane anchoring sequence; antibody;
KW	vaccine; immunotherapy;	malaria; ds.
XX		
OS	Plasmodium falciparum.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	77..5605
FT	ET	/*tag= a
FT	ET	/product= liver stage antigen-3
FT	ET	77..274
FT	exon	

in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of *P. acnes* in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for *P. acnes* proteins. These antibodies can be used to downregulate expression and activity of *P. acnes* polypeptides and therefore treat *P. acnes* infections. The antibodies may also be used as diagnostic agents for determining *P. acnes* presence, for example, by enzyme linked immunosorbent assay (ELISA). This sequence encodes the polypeptides shown in AAU6082-AAU6099 and AAU67848.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.int/pub/published_pct_sequences.

Sequence 1473 BP; 299 A; 390 C; 482 G; 299 T; 3 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	0.5	1473	47	30	70	37
Percent Similarity:	91.50					
Best Local Similarity:	41.85%					
Query Match:	8.57%					
DB:	23					

US-10-024-955-7 (1-213) x AAS59711 (1-1473)

QY 26 ThrGluGluIleAsnLysAlaIleAspAlaIleAlaIleGluInserGlu--- 44

DB 665 ACTGAAAGATTTGGAAACAGGAGATGCTCCAGACCCAGCAGCAATCCGAGATC 724

QY 45 ---ThrIleAspProMetLysValProAspHisAlaAsp-----LysPhe 58

DB 725 GGGGCGCTTGATCTCTATGCT-----GATGATGGATCATCTGCGCATCCCACTTC 778

QY 59 GluArgHisValGlyIleValAspPheLysGly-----GluLeuAlaMetArgAsn 75

DB 779 GCTGTGAGCGCTGGAGATTTCCGATGTTCCGATCATATGAGAGCGCGACTCGATTC 838

QY 76 IleGluAlaArg-----GlyLeuLysGlnMetLysArgGlnGlyAsp 89

DB 839 GTTGAGAGAGAGAGCGCAATCTTCCGACCCCTGACGATCTTCCCGCTGGGTGAT 898

QY 90 AlaAsnValLysGlyGluGluIleValLysAlaHisLeuLeuIleGlyValHisAsp 109

DB 899 -----CTGAGCGGCTGAGAG-----CTCAGAGCACTTGCCTGGCCGCTTTCGAC 946

QY 110 AspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThr 129

DB 947 GATCTCGTGGCTCCAGAGCGGCTGCTTGTGGACATGCGCCAGATTAATGCTGCGCC 1006

QY 130 HisValIleSerAspIleGlnAspPheValAla----- 141

DB 1007 CCAGGACACTGAGCGCTTACG-----GTGGTCACCCAACTCTCTACTTCCGAACCT 1060

QY 142 LeuSerLeuGluIleSerAspGluGly-----AsnIleThrMetThrSerPhe 157

DB 1061 ACTCAGATGATTAATGCTGCTGACGACCTCCGCGCTACCGGATTAATGACGCTGATCAC 1120

QY 158 GluValArgGlnPheAlaAsnValValAsnHisIleGlyIleLeuSerIleLeuAspPro 177

DB 1121 CCGGTCGCGCAGGTACGAGATGCCCTTATTCGACGGGCTGTGCTCAACTGC---GACCGG 1177

QY 178 IlePheGlyVal 181

DB 1178 TTAGATGTGT 1189

RESULT 12

AAA70259 standard: DNA; 4677 BP.

AC AAA70259;

XX 07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:392.

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacide; infection; insecticide; ds.

OS Plasmodium falciparum.

PN WO200025728-A2.

PD 11-MAY-2000.

PF 05-NOV-1999; 99MO-US26796.

PR 05-NOV-1998; 98US-0107131.

PA (HOFF/) HOFFMAN S.

PA (CARU/) CARUCCI D.

PA (GARD/) GARDNER M.

PI (VENT/) VENTER J C.

DR Hoffman S, Carnucci D, Gardner M, Venter JC;

WPI; 2000-365347/31.

PT Proteins encoded by chromosome 2 of the human malarial parasite,

PT Plasmodium falciparum, useful as antimalarial vaccines and in the

PT diagnosis of *P. falciparum* infection.

PS Disclosure; Page 565-566; 577pp; English.

CC The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*.

CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2) vaccines against *P. falciparum* infection comprising (I) or (II).

CC (I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are

CC useful in the detection of infection with *P. falciparum*. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in

CC *P. falciparum*. Sequencing of the *Plasmodium* chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the

CC complexity of the parasite lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many

CC parts of the world, and there is a pressing need for vaccines and new drugs. AAA70078 to AAA70287 and AAB10144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

CC Sequence 4677 BP; 2106 A; 402 C; 966 G; 1203 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Oy	151	---asnlterhmettrserphiegulvalarglnphlealsnvalasnhisille	169
Db	1078	AACAATATTTGAANAAGACGTAACTAGTACAAAAGACTAGTTTAAACAACTAAAGAGAAA	1137
Oy	170	glycylleuserile	174
Db	1138	TTTGACTAGCTGTA	1152
		RESULT 9	
		AAH41223/C	
ID	AAH41223	standard; DNA; 349980 BP.	
XX	AAH41223;		
AC			
XX			
DT	29-OCT-2001	(first entry)	
XX			
DE	Pyrococcus abyssi genomic fragment #2.		
XX			
KW	Hyperthermophilic archaeon; hyperthermophilic protein; ds.		
XX			
OS	Pyrococcus abyssi.		
XX			
FT	Key	Location/Qualifiers	
FT	misc_feature	1..49980	
FT		/tag= "a	
FT		/note="This sequence overlaps with the 3' end of	
FT		AAH86431"	
FT	misc_feature	300001..349980	
FT		/tag= "b	
FT		/note="This sequence overlaps with the 5' end of	
FT		AAH41224"	
XX			
FM	FR2792651-A1.		
XX			
PD	27-OCT-2000.		
XX			
PE	21-APR-1999;	99FR-0005034.	
XX			
PR	21-APR-1999;	99FR-0005034.	
XX			
PA	(CNRS) CNRS CENT NAT RECH SCI.		
PA	(IFRE-) IFREMER INST FR RECH EXPL MER.		
XX			
PI	Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;		
PI	Querellou J, Weissenbach J, Saurin W, Hellig R;		
XX			
DR	WPI; 2001-126236/14.		
XX			
PT	New nucleotide sequences isolated from Pyrococcus abyssi encode		
PT	proteins useful in industry -		
XX			
PS	Claim 1; Page 265-361; 1657pp; French.		
XX			
CC	The present invention relates to the genomic sequence of Pyrococcus		
CC	abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a		
CC	hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal		
CC	vents. The present sequence is a fragment of the genomic sequence of P.		
CC	abyssi. The 5' end of this sequence overlaps with the 3' end of AAB86431		
CC	and the 3' end of this sequence overlaps with the 5' end of AAB41224. The		
CC	proteins of the present invention have various potential industrial uses,		
CC	since the proteins are stable at very high temperatures, some up to 110		
CC	degrees centigrade.		
CC	Note: This patent is in the same patent family as WO200065062, which		
CC	contains additional sequences as shown in AAB99132-AAB99143,		
CC	AAH75903-AAH75920 and AAG6436.		
XX			
SQ	Sequence 349980 BP; 92953 A; 77841 C; 81831 G; 97355 T; 0 other;		
	Alignment Scores:		
	Pred. No.:	517	Length: 349980
	Score:	93.50	Matches: 60
	Percent Similarity:	37.11%	Conservative: 35

Oy 151 ---AsnIleThreMetHsrSerPheGluValArgGlnPheAlaAsnValValaAsnHisIle 169
DB 1078 AACATATTGAATGACAGCTGTAATGACAAAGACTATGTAACACTTAAGAGAAA 1137
Oy 170 GlyIleuSerIle 174
DB 1138 TTTGGACTAGCTGTGA 1152
RESULT 8
ABQ72644
ID ABQ72644 standard; cDNA: 1233 BP.
AC ABQ72644;
XX
DT 03-SEP-2002 (first entry)
XX
DE Human MDR1 encoding cDNA SEQ ID NO 196.
XX
KW Human; MDR1; disease detection and treatment molecule polynucleotide;
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antiout;
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200240715-A2.
XX
PD 23-MAY-2002.
XX
PF 06-SEP-2001; 2001WO-US27628.
XX
PR 06-SEP-2000; 2000US-230505P.
PR 06-SEP-2000; 2000US-230514P.
PR 06-SEP-2000; 2000US-230515P.
PR 06-SEP-2000; 2000US-230517P.
PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000US-230519P.
PR 06-SEP-2000; 2000US-230595P.
PR 06-SEP-2000; 2000US-230597P.
PR 06-SEP-2000; 2000US-230598P.
PR 06-SEP-2000; 2000US-230599P.
PR 06-SEP-2000; 2000US-230610P.
PR 06-SEP-2000; 2000US-230655P.
PR 06-SEP-2000; 2000US-230888P.
PR 06-SEP-2000; 2000US-230989P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231167P.
XX
PA (INCYTE) INCYTE GENOMICS INC.
XX
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstin EH, Peralta CH, David WH, Panzer SR, Flores V, Daffo A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
XX
DR WPI; 2002-527544/56.
DR P-PSDB; ABP51427.
XX
XX Novel human disease detection and treatment polypeptide, useful in
PI diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
PT e.g. AIDS
XX
PS Claim 1; Page 411; 618pp; English.
XX
CC The invention relates to an isolated human disease detection and
CC treatment (MDR1) polypeptide (I) selected from a polypeptide having a

CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound for
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
CC screening a compound for effectiveness in altering expression of a target
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDR1 in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDR1 in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDR1.
CC (I) or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDR1, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases, in somatic or
CC germ-line gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences.
XX
XX SQ Sequence 1233 BP; 489 A; 184 C; 256 G; 304 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0.114 Length: 1233
XX Score: 96.00 Matches: 44
XX Percent Similarity: 40.54% Conservative: 31
XX Best Local Similarity: 23.78% Mismatches: 58
XX Query Match: 8.99% Indels: 52
XX DB: 24 Gaps: 10
XX
XX US-10-024-955-7 (1-213) x ABQ72644 (1-1233)
Oy 27 GIuGIuIleAsnLysAlaIleAspAspAlaIleAlaIleGluGlnSerGluThrIle 46
DB 643 GAATACATCAACAGAAAGTTTGAATGATTAGTAAGAAGTTAAAGACAGCGTTGAA 702
Oy 47 Asp---PrometLysVal-----ProAspHisAlaAspLysPheGlu 59
DB 703 AATGGTCATCTTCAGTGGTCACAAAGATCTTAATGTCACATGCTGCATATGTC--- 759
Oy 60 ArgHisValGlyIleValAspPhe-----LysGlyGlu--- 70
DB 760 CGCCATTTTCACCAATATCATGTTTATATACATGACACATAGCAGCAAGCAAGCAAG 819
Oy 71 -----LeuAlaMetArgAsnIleGluAlaArgly----- 80
DB 820 TGCAGACATTTATTTATTCAGCTGCAGTGAAGTGGTGAAAGAGATGTCGAAAGAT 879
Oy 81 LeuLysGlnMetLysArgGlnLysAspAlaAsnValLysGlyGlu----- 96
DB 880 GTAAATGATATGCTTCATCAACAGAAATGCGAAGTTTAAATGAAGCTTACTTACAGAA 939
Oy 97 ---GlyIleValLysAlaHisLeu-----LeuIleGlyValHisAspIleValSer 113
DB 940 ATCAATATTCACATAAGCAGCACTTTCAGAGATTGAACAGACAGACAGAACTTTAT 999
Oy 114 MetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrHisValIleSer 133
DB 1000 TTGAGG-----AACCAATAAAG 1017
Oy 134 AspIleGlnAspPheValValAlaLeuSerLeuGluIleSerAspGlyGly----- 150
DB 1018 GATTTAAGGAGATCTTTTCATTCAGATATCTTTTATAGTAGAGAGACAGAGAGATC 1077

```
Db      806 TTGGACTAGCTGTA 820
|||||.....
RESULT 7
AAS31185
ID      AAS31185 standard; cDNA: 1233 BP.
XX
XX      AAS31185;
XX
XX      04-DEC-2001 (first entry)
XX
XX      Human diagnostic and therapeutic polynucleotide (DITHP) #200.
DE
XX
XX      Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW      cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW      acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW      respiratory disorder; ss.
XX
XX      Homo sapiens.
OS
XX
XX      MO200162927-A2.
PN
XX
XX      30-AUG-2001.
PD
XX
XX      21-FEB-2001; 2001WO-US06059.
PF
XX
XX      24-FEB-2000; 2000US-0184693.
XX      24-FEB-2000; 2000US-0184697.
XX      24-FEB-2000; 2000US-0184698.
XX      24-FEB-2000; 2000US-0184768.
XX      24-FEB-2000; 2000US-0184769.
XX      24-FEB-2000; 2000US-0184770.
XX      24-FEB-2000; 2000US-0184771.
XX      24-FEB-2000; 2000US-0184772.
XX      24-FEB-2000; 2000US-0184773.
XX      24-FEB-2000; 2000US-0184774.
XX      24-FEB-2000; 2000US-0184776.
XX      24-FEB-2000; 2000US-0184777.
XX      24-FEB-2000; 2000US-0184797.
XX      24-FEB-2000; 2000US-0184813.
XX      24-FEB-2000; 2000US-0184837.
XX      24-FEB-2000; 2000US-0184841.
XX      24-FEB-2000; 2000US-0185213.
XX      24-FEB-2000; 2000US-0185216.
XX      12-MAY-2000; 2000US-0203785.
XX      15-MAY-2000; 2000US-0204226.
XX      16-MAY-2000; 2000US-0204525.
XX      16-MAY-2000; 2000US-0204821.
XX      16-MAY-2000; 2000US-0204908.
XX      16-MAY-2000; 2000US-0205232.
XX      17-MAY-2000; 2000US-0204815.
XX      17-MAY-2000; 2000US-0204863.
XX      17-MAY-2000; 2000US-0205221.
XX      17-MAY-2000; 2000US-0205285.
XX      17-MAY-2000; 2000US-0205286.
XX      17-MAY-2000; 2000US-0205287.
XX      17-MAY-2000; 2000US-0205323.
XX      17-MAY-2000; 2000US-0205324.
XX
XX      (INCY-) INCYTE GENOMICS INC.
PA
XX      Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI      Chen A, D'sa SA, Amsley S, Dahl CR, Dam TC, Daniels SE;
PI      Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
PI      Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI      Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI      Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
XX      WPI: 2001-502867/55.
DR      P-PSDB; AAU19614.
PT      Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
        enzymes, hormones and receptors, useful in diagnostics and therapeutics
```

```
PT
XX
XX      Claim 1; Page 392; 522pp; English.
XX
XX      The invention relates to polynucleotides (I) encoding diagnostic and
CC      therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC      and proteins involved in growth and development and receptors. (I) and
CC      (II) may be used in the prevention, diagnosis and treatment of diseases
CC      associated with inappropriate DITHP expression. For example, (I) and
CC      (II) may be used to treat disorders associated with decreased polypeptide
CC      expression by rectifying mutations or deletions in a patient's genome,
CC      that affect the activity of the DITHPs, by expressing inactive proteins
CC      or supplementing the patient's own production of them. (I) and (II)
CC      may be used to treat diseases, for example, cell proliferative disorder,
CC      Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC      leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC      (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC      into a host cell and culturing the cell to express the protein. (I) and
CC      its complementary sequences may also be used as DNA probes in diagnostic
CC      assays to detect and quantitate the presence of similar nucleic acids in
CC      samples, and therefore which patients may be in need of restorative
CC      therapy. (II) may also be used as antigens in the production of
CC      antibodies against DITHPs and in assays to identify modulators of DITHP
CC      expression and activity. The anti-DITHP antibodies and antagonists may
CC      also be used to down regulate expression and activity. The anti-DITHP
CC      antibodies may also be used as diagnostic agents for detecting the
CC      presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
CC      assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and
CC      therapeutic (DITHP) polynucleotides of the invention.
XX
XX      Sequence 1233 BP; 489 A; 184 C; 256 G; 304 T; 0 other;
SO
XX
XX      Alignment Scores:
XX
XX      Pred. No.:      0 114      Length:      1233
XX      Score:          96.00      Matches:      44
XX      Percent Similarity: 40.54%      Conservative: 31
XX      Best Local Similarity: 23.78%      Mismatches: 58
XX      Query Match:      8.99%      Indels:      52
XX      DB:              22      Gaps:      10
XX
XX      US-10-024-955-7 (1-213) x AAS31185 (1-1233)
OY      27 GUGUUIleasnlyalaileaspaspalailealailegluglnserclurhile 46
DB      643 GAATACATCAACAGAAAGTTGATGATTAGTTAAAGAACTTAAGAGAGGTTGAA 702
OY      47 Asp---Prometlylval-----Proasphialaspyspbeclu 59
DB      703 AATGGTCATCTTCAGTGGTCAAGATACTTAATCTCAGACATGCTTCATGTTTC--- 759
OY      60 ArgHisvalGlyIlevalAspHe-----LysGlyIgu--- 70
DB      760 GGCATTTTCAGCAAAATCATGTTTATATACATGACACAAATGACGAAAGCAAGAGAG 819
OY      71 -----LeuAlaMetArgasnilegluaAlaArgly----- 80
DB      820 TGCAAGACATTATTATTACGTCAGCTGCAAGTGTGGAAGAAAGATGTGAAAGACAT 879
OY      81 LeuLysGlnMetLysArgGlnGlyaspAlaasnValLysGlyGluIu----- 96
DB      880 GTAATGATATGCTTCATCAACGAAGAAATGGCAAGCTTTTAATGAAGCTTACTTACAGAA 939
OY      97 ---GlyIleValLysAlaHisLeu-----LeuIleGlyValHisAspAspIleValser 113
DB      940 ATCAATATTCACCTAAAGACAACTTTCAGAGATGAACAGACAGCAAGAACTGTATAT 999
OY      114 MetGluTyrAspLeuAlaTylLysLeuGlyAspLeuHisProthThHisValIleSer 133
DB      1000 TTGGAG-----AACCAATAAAG 1017
OY      134 AspIleGlnAspPheValValaLeuSerLeuGlnIleSerAspGluIy----- 150
DB      1018 GATTTAAGGAGATCTTTCATTCAGATATCTGTTTAAAGAGAAACAGAGAGAGCATC 1077
```



```

Db 547 GTAATGATATGCTTCATCAAGAAATGGAGTTTAAATGAAGCTTACTTACAGAA 606
QY 97 ---GlytIleValLysAlaHisIleu-----LeuIleGlyValHisAspAspIleValSer 113
Db 607 ATCAATATCAGTAAAGCAACACTTTCAGAGATTGAACAGACAGACAGAACTGTATAT 666
QY 114 MetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThHisValIleSer 133
Db 667 TTGGAG-----AACCAATATAAG 684
QY 134 AspIleGlnAspPheValValAlaLeuSerLeuGlnIleSerAspGlyGly-----150
Db 685 GATTAAAGGAGCTTTTCATTCAGATATCTTTAGTAGAGAGAAACAAGAGAGACATC 744
QY 151 ---AsnIleThrMetThrSerPheGluValArgGlnPheAlaAsnValValAsnHisIle 169
Db 745 AACATATATTGAATGACAGTGAATATGTAACAAGAATGTTTACCAATACTTAAGAGAA 804
QY 170 GlyGlyLeuSerIle 174
Db 805 TTGGAGCTAGCTGTA 819
RESULT 5
ABA03885
ID ABA03885 standard; cDNA; 892 BP.
XX
AC ABA03885:
XX
DT 14-FEB-2002 (first entry)
XX
XX Human POLY16 nucleotide sequence SEQ ID NO:31.
XX
KW Human: POLYX; gamma aminobutyric acid receptor; GABA receptor;
KW epidermal growth factor; EGF; complement receptor; HSPC; syntaxin;
KW haematopoietic stem and progenitor cell; sulphotransferase; proliditin;
KW antidepressant; cerebroprotective; antiparkinsonian; nootropic; relaxant;
KW anticonvulsant; neuroleptic; neuroprotective; antialcoholic; cardiant;
KW tranquilliser; antiarrhythmic; psychiatric; medical; depression; stroke;
KW Parkinson's disease; Huntington's disease; Tourette's syndrome; anxiety;
KW amyotrophic lateral sclerosis; head trauma; Alzheimer's disease;
KW alcoholism; vigilance; muscle tension; epileptogenic; memory function;
KW cardiomyopathy; arrhythmogenic right ventricular dysplasia; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 4..888
XX FT /*tag= a
XX FT /product= "POLY16"
XX
PN WO200179294-A2.
XX
PD 25-OCT-2001.
XX
PF 19-APR-2001; 2001WO-US12854.
XX
PR 19-APR-2000; 2000US-198293P.
PR 20-APR-2000; 2000US-198645P.
PR 25-APR-2000; 2000US-199476P.
PR 26-APR-2000; 2000US-199880P.
PR 26-APR-2000; 2000US-200024P.
PR 26-APR-2000; 2000US-200025P.
PR 09-JUN-2000; 2000US-210809P.
PR 17-JUL-2000; 2000US-218591P.
PR 11-AUG-2000; 2000US-224610P.
PR 09-FEB-2001; 2001US-267673P.
XX
XX (CURAGEN CORP.
XX
XX Taupier RJ, Vernet CM, Fernandes E, Shinkets RA, Majumder K;
PI Padigaru M, Colman SD, Zernhusen BD, Splytek KA, Burgess CE, Liu X;
XX WPI; 2002-017601/02.

```

```

DR P-PSDB: AAG68269.
XX
XX New isolated polypeptides for treating a broad range of pathological
PT states, e.g., depression, stroke, Parkinson's disease, Huntington's
PT disease, Tourette's syndrome, amyotrophic lateral sclerosis, head
PT trauma, and Alzheimer's
XX
XX Claim 9; Page 61; 155pp; English.
XX
XX The present invention describes polypeptides (I), designated POLYX
CC polypeptides (e.g. POLY1, POLY2, etc.), and the polynucleotide
CC sequences (II) encoding them. POLY1-4 are members of the gamma
CC aminobutyric acid (GABA) receptor family; POLY5-8 are members of the
CC epidermal growth factor (EGF) family; POLY9-11 are members of the
CC complement receptor family; POLY12 is a member of the haematopoietic
CC stem and progenitor cell (HSPC) family; POLY13 is a member of the
CC sulphotransferase family; POLY14-16 are members of the syntaxin family;
CC and POLY17 is a member of the proliditin family. (I) and (II) can have
CC antidepressant, cerebroprotective, antiparkinsonian, nootropic, relaxant,
CC anticonvulsant, neuroleptic, neuroprotective, antialcoholic, cardiant,
CC tranquilliser and antiarrhythmic activities. (I) and (II) can be used
CC for treating or preventing a POLYX-associated disorder in humans as a
CC therapeutic in the manufacture of a medicament for treating a syndrome
CC associated with a human disease selected from a POLYX-associated
CC disorder, for treating a pathological state in a mammal, especially
CC patients suffering from, e.g., psychiatric and medical conditions,
CC depression, stroke, Parkinson's disease, Huntington's disease,
CC Tourette's syndrome, amyotrophic lateral sclerosis, head trauma,
CC Alzheimer's disease, alcoholism, vigilance, anxiety, muscle tension,
CC epileptogenic activity and memory functions, cardiomyopathy and
CC arrhythmogenic right ventricular dysplasia. The present sequence
CC encodes POLY16.
XX
XX Sequence 892 BP; 355 A; 130 C; 185 G; 222 T; 0 other;
S0
Alignment Scores:
Pred. No.: 0.0729 Length: 892
Score: 96.00 Matches: 44
Percent Similarity: 40.54% Conservative: 31
Best Local Similarity: 23.78% Mismatches: 58
Query Match: 8.99% Indels: 52
DB: 24 Gaps: 10
US-10-024-955-7 (1-213) x ABA03885 (1-892)
QY 27 GUGUUIleAsnLysAlaHisIleu-----LeuIleGlyValHisAspAspIleValSer 46
Db 310 GAATACATCAACAGAAAGTTTGAAATGATTGTTAAAGAGTTTAAAGAGGTTGAA 369
QY 47 Asp---PrometLysVal-----ProAspHisAlaAspLysPheGlu 59
Db 370 AATGGTCCATCTTCAGTGGTGCACAAAGATCTTAATCTCAGACATGCTGCAATGTC--- 426
QY 60 ArgHisValGlyIleValAspPhe-----LysGlyGlu---70
Db 427 CGCCATTTTTCAGCAAAATGATGTTTATATACATGACACAAATGACGAAAGCAAGAGAA 486
QY 71 -----LeuAlaMetArgAsnIleGlyAlaValGly-----80
Db 487 TGCAGACATTTATTTTACGTCAGCTGGAAGTTGCTGGAAGAGATGTCGAGAGAAGAT 546
QY 81 LeuLysGlnMetLysArgGlnGlyAspAlaAsnValLysGlyGlu-----96
Db 547 GTAATGATATGCTTCATCAAGAAATGGAGTTTAAATGAAGCTTACTTACAGAA 606
QY 97 ---GlytIleValLysAlaHisIleu-----LeuIleGlyValHisAspAspIleValSer 113
Db 607 ATCAATATCAGTAAAGCAACACTTTCAGAGATTGAACAGACAGACAGAACTGTATAT 666
QY 114 MetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrThHisValIleSer 133
Db 667 TTGGAG-----AACCAATATAAG 684

```



```

FT      /tag= b
FT      /note= "Df1 primer"
FT      misc_feature
FT      247..263
FT      /tag= c
FT      /note= "Df4 primer"
FT      559..582
FT      /tag= d
FT      /note= "Df5 primer"
FT      684..664
FT      /tag= e
FT      /note= "Df6 primer"
FT      225..208
FT      /tag= f
FT      /note= "Df8 primer"

W09420614-A.
PD      15-SEP-1994.
PF      11-MAR-1994; 94WO-AU00117.
PR      12-MAR-1993; 93US-0031141.
PR      22-JUN-1993; 93US-0081540.
XX      (CHIL-) INST CHILD HEALTH RES.
XX      Chua K, Thomas WR;
XX      MPI: 1994-303021/37.
XX      P-PSDB: AAR60576.
DR      New nucleic acid encoding specific dust mite allergens - and
XX      related vectors, transformed cells, peptides and antibodies,
XX      useful for desensitisation and diagnosis.
XX      Claim 13; Page 38-39; 67pp; English.
XX      DerfVII antigen is useful as anti-allergic reagent for treating
XX      CC sensitivity to house dust mite allergens. The DNA can be used
XX      as a probe to detect the sensitivity of an individual to the
XX      allergen.
SQ      Sequence 761 BP; 245 A; 134 C; 137 G; 245 T; 0 other;

Alignment Scores:
Pred. No.: 3 53e-118 Length: 761
Score: 1068.00 Matches: 213
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-024-955-7 (1-213) x AAO71401 (1-761)
OY      1 MetMetLysPheLeuLeuIleAlaValAlaPheValAlaValSerAlaAspProIle 20
DB      43 ATGAGCAATTTTGTGATGTCGCGCATTTGCGCGTTCGCGTATTCATTT 102
OY      21 H1STYFASPLYSIIETHRGLUGLULIleasnlysaIleAspSPAlaIleAlaIle 40
DB      103 CACTATGATAAATCACCCAGAAATCAACAAGCATATGATGATCCATTCCTATT 162
OY      41 GIUGInserGluThrIleAspPrometLysValProAspHisAlaAspLysPheGluArg 60
DB      163 GAAACAATCCGAACAATATCAATGAAAGTACCTGATCAAGCGATAAATTCGAACT 222
OY      61 HisValGlyIleValAlaSPheLysGlyLulLeuAlaMetArgAsnIleGluAlaArgGly 80
DB      223 CATGTTGATTTGTGATTCGAAGGTGAATTAAGCATTCGAACAATTCAGGCTGAGGA 282
OY      81 LeuLysGlnMetLysArgGlnGlnLysPalaAsnValLysGlyGlnGlnGlyIleValLys 100
DB      283 TTGAACAACAATGAACGTCACAGGTGATGCTAATGTCACAAAGTGAGAGGATATTGTTAA 342

```

```

OY      101 AlaHisLeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
DB      343 GCTCATTTGTGATGCGGTGTCACGATGATATGCTCCATGATGAAATATGATTTAGCATAC 402
OY      121 LysLeuGlyAspLeuHisProThrThrHisValIleSerAspIleGlnAspPheVal 140
DB      403 AAATGGGTGATCTTCATCCAACTCAATGATTCGATTCGATTCGAATTCGAATTTGTTGTT 462
OY      141 AlaLeuSerLeuGlnIleSerAspGlnGlyAsnIleThrMetThrSerPheGluValArg 160
DB      463 GCCTTGCCCTGAAATTCGATGAAGTAACATGAATGACATCTTTTGAAGTACGA 522
OY      161 GlnPheAlaAsnValAlaAsnHisIleGlyLysSerIleLeuSerIleLeuAspProIlePheGly 180
DB      523 CAATTCGATTAATGTTTCACACCATATGATGGTGGTCTTTCAATCTTGATCCAAATTTTGGC 582
OY      181 ValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArgLysGlnMetThrLys 200
DB      583 GTTTATCTGATGATATGACCGCTATTTCCAAAGACACCGTACGTAAGAAATGACCAAA 642
OY      201 ValLeuAlaProAlaPheLysArgGlnLulLeuGlnLulLysAsn 213
DB      643 GTAATGGACACACCATTTAAACGTGAATTGAAAAAAAAT 681

RESULT 2
AAO71400
ID      AAO71400 standard; cDNA; 812 BP.
XX      AAO71400;
XX      01-APR-1995 (first entry)
DE      House dust mite allergen DerpVII cDNA.
XX      DerpVII allergen; anti-allergic; allergy diagnosis; ss.
XX      Dermatophagoides pteronyssinus.
XX      Key Location/Qualifiers
FT      CDS 68..712
FT      /tag= a
FT      misc_feature 119..136
FT      /tag= b
FT      /label= primer
FT      /note= "SEQ ID NO:3"
FT      misc_feature 272..288
FT      /tag= c
FT      /label= primer
FT      /note= "SEQ ID NO:4"
FT      misc_feature 584..607
FT      /tag= d
FT      /label= primer
FT      /note= "SEQ ID NO:5"

W09420614-A.
PD      15-SEP-1994.
PF      11-MAR-1994; 94WO-AU00117.
PR      12-MAR-1993; 93US-0031141.
PR      22-JUN-1993; 93US-0081540.
XX      (CHIL-) INST CHILD HEALTH RES.
XX      Chua K, Thomas WR;
XX      MPI: 1994-303021/37.
XX      P-PSDB: AAR60575.
PT      New nucleic acid encoding specific dust mite allergens - and
PT      related vectors, transformed cells, peptides and antibodies,

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 20, 2003, 14:31:27 ; Search time 236 Seconds
(without alignments)
2032.524 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 1068
Sequence: 1 MKKFLLIAAVFAVAVSADPI.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=+lp
-Q=/cgn2_1/USPTO_SPOOL/US10024953/runat_14022003_130657_23695/app_query.fasta_1.391
-DB=N_Geneseq_101002 -OPMT=fastap -SUFFIX=ing -MIMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=50
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10024955.ecgn_1_1.396.&runat_14022003_130657_23695 -NCPU=6 -ICPU=3
-NO_XLPLY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARRN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1068	100.0	761	15	AAO71401
2	959	89.8	812	15	AAO71400
3	99.5	9.3	826	23	ABL14609
4	96	9.0	892	24	ABA03884
5	96	9.0	892	24	ABA03885
6	96	9.0	893	24	ABA03883
7	96	9.0	1233	22	AAS31185
8	96	9.0	1233	24	ABO72644
9	93.5	8.8	349980	22	AAH41223
10	93	8.7	1308	24	ABN68457
11	91.5	8.6	1473	23	AAS59711
12	90	8.4	4677	21	AAH70259
13	90	8.4	5361	18	AAH78868
14	90	8.4	5529	24	ABK50883
15	90	8.4	6152	18	AAH78867
16	89	8.3	3496	23	AAS89799
17	89	8.3	5066	19	AAV52212
18	88.5	8.3	1131	14	AAQ46032
19	88.5	8.3	1903	21	AAF21640
20	88.5	8.3	2156	24	ABK84110
21	88.5	8.3	2157	19	AAV24010
22	88	8.2	1311	21	AAH54516
23	88	8.2	1311	21	AAZ91826
24	88	8.2	2700	23	AAS93828
25	87.5	8.2	349980	22	AAH41225
26	87	8.1	1308	24	ABN68458
27	86.5	8.1	1980	23	AAS51833
28	86.5	8.1	1995	23	AAS54444
29	86	8.1	934	22	AAH53069
30	86	8.1	2609	22	AAH02108
31	86	8.1	4411529	22	AAI99682
32	85.5	8.0	1693	21	AAH75637
33	85.5	8.0	2832	23	AAS54100
34	85.5	8.0	2832	23	AAH54100
35	85.5	8.0	5688	24	ABK14973
36	85.5	8.0	9436	22	AAH16294
37	85.5	8.0	9437	22	AAH16286
38	85.5	8.0	58407	19	AAV21210
39	85.5	8.0	1664976	19	AAV21209
40	85	8.0	2158	23	ABL15311
41	85	8.0	3277	23	ABL21690
42	85	8.0	4158	23	ABL15310
43	84.5	7.9	2454	22	AAH52328
44	84.5	7.9	2454	22	AAH53067
45	84.5	7.9	2472	24	ABN91781

ALIGNMENTS

RESULT 1
AAO71401
ID AAO71401 standard; cDNA; 761 BP.
XX
AC AAO71401;
XX
DT 01-APR-1995 (first entry)
XX
DE House dust mite allergen DerfVII CDNA.
XX
KW DerfVII allergen; antiallergic; allergy diagnosis; ss.
XX
OS Dermatophagoides farinae.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT msc_feature 94..111

Percent Similarity: 46.858
Best Local Similarity: 31.47%
Query Match: 8.47%
DB: 9
Conservative: 22
Mismatches: 40
Indels: 36
Gaps: 7

US-10-024-955-7 (1-213) x AC098483 (1-177865)

```
QY 74 ArgAsnIleGluAlaArgGlyLeuLysGlnMetLysArgGlnGlyAspAlaAsnValLys 93
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28546 CAAACCTTGATATGCGCTGGCGCTATTTGACCATAAACAGCATATATAGCCCTAAATAGG 28605
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 94 GlyGluGlyIleValLysAlaHisLeu---LeuIleGlyValHisAspAspIleVal 112
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28606 GGGTCT-----TCTGCCAAGCTGGGTGTGTAGGAAACATGGCAATATTCMA 28653
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 113 SerMetGluTyrAspLeuAlaTyrLysLeuGlyAspLeuHisProThrHisValIle 132
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28654 AATGGCTGCAGACCTCCAGTTCCAAAGTGAAGATACCATGCC----- 28698
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 133 SerAspIleGlnAspPheValAlaLeuSerLeuGluIleSerAspGluGlyAsn--- 151
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28699 -----ATCCAGACATTCGAGTTGCCCTTCCCTGATGATAGTGATGTGAGCACTT 28752
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 152 -----IleThrMetThrSerPheGluVal-----ArgGlnPhe 162
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28753 GTCATATACGTGCTAGCCATTTGTATGTCTTTTGAATAATATGTTTTCGATGCTTTT 28812
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 163 AlaAsn-----ValAlaAsnHisIleGlyGly 171
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28813 GCCCATTTTAAATCAATTTATTGTGAGATATTTTGGTACTCAGTTACTTGAGTCT 28872
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 172 LeuSerIleLeuAspProIlePheGlyValLeuSerAspValLeuThrAla-IlePheG 191
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28873 TTATATATTCTCGAT-----ATGAACCCCTTGTCAATATGATGTTGGCAATCTTTC 28926
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

QY 191 nAspThr 193
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 28927 TCCCACT 28933
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

Search completed: February 20, 2003, 17:13:48
Job time : 2574 secs

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

BglII

EcoRI

HindIII

SeqDerMap FngPrnt

SeqDerMap FngPrnt

SeqDerMap FngPrnt

SeqDerMap FngPrnt

6622	6388	8696	8822	5834	5709
2067	2146	6	<800	6382	6463
4428	4673	1147	1143	512	<800
5977	5833	3738	3635	449	<800
5951	5833	105	<800	2030	1979
2648	2581	802	809	3178	3301
961	965	2806	2887	643	<800
5180	4962	1166	1143	9246	9131
2464	2581	1124	1143	282	<800
5503	5833	4962	4914	2919	2952
2140	2146	3943	4067	2263	2274
4447	4223	2200	2191	3574	3619
847	850	4168	4067	35	<800
2229	2325	2342	2374	6436	6463
1045	1053	6480	6501	4700	4706
1128	1120	1721	1744	588	<800
2373	2325	8305	8253	4579	4512
352	<800	518	<800	4394	4364
202	<800	883	909	1662	1608
2273	2325	459	<800	1627	1608
557	<800	1085	1143	1333	1315
592	<800	556	<800	6089	6077
3400	3336	1527	1481	4162	4161
1324	1278	535	<800	490	<800

1234	1182	1275	1250	589	<800
224	<800	2556	2550	21	<800
1658	1591	555	<800	7711	7778
956	965	578	<800	257	<800
2571	2581	2294	2374	1204	1169
1143	1120	1483	1481	225	<800
1085	1120	2091	2191	2873	2952
9634	9789	572	<800	2727	2717
4054	4010	32	<800	2661	2717
1071	1120	1137	1143	3011	2952
836	850	4208	4067	2391	2483
121	<800	1474	1481	604	<800
1296	1278	2657	2675	5034	4955
4247	4223	443	<800	6444	6463
1108	1120	5072	4914	607	<800
1592	1591	1523	1481	20573	20626
438	<800	4023	4067	1774	1781
1205	1120	3365	3334	2328	2483
1906	1960	1165	1143	1801	1781
1971	1960	3819	3635	3035	2952
4407	4673	1931	1968	3919	3929
5464	5353	414	<800	3948	3929
313	<800	3527	3635	259	<800
4833	4962	1196	1250	266	<800
3392	3336	8929	8822	2504	2483
4784	4962	202	<800	3248	3301
5187	5353	7104	7041	8210	8181
8773	8785	437	<800	2253	2274
2355	2325	1173	1143	1510	1480
1294	1278	3570	3635	40	<800
2349	2325	8637	8822	1172	1169
2218	2325	2906	2887	379	<800
1017	1053	1180	1143	323	<800
2120	2146	931	909	596	<800

Alignment Scores:
Pred. No.:
Score:

1.53e+04
90.50

Length:
Matches:

177865
45

```

repeat_region      /rpt_family="AT_rich"
                    28332
repeat_region      /rpt_family="AT_rich"
                    complement(29191..29216)
gene               /rpt_family="AT_rich"
                    29728..30324
                    /gene="At2g15010"
                    /note="T15J14.5"
                    /join(29728..29976,30060..30137,30227..30324)
                    /gene="At2g15010"
                    /join(29745..29976,30060..30137,30227..30324)
                    /gene="At2g15010"
                    /codon_start=1
                    /product="putative thionin"
                    /protein_id="AAD03358.1"
                    /db_xref="GI:4115356"
                    /translation="MEKTYIFSVLVWGLYISQIOVEAOKTCPCPSOSTRKEPFCISE
                    GNIQILCSAESGCRDITVGYCPGSPFGLTNSGDVNVYCKLQCVSSLCGALTSLQK
                    LDYSGKVAVERCTKACSTICTKSKTAETV"
                    30796..30837
repeat_region      /rpt_family="AT_rich"
                    complement(36100..36131)
repeat_region      /rpt_family="AT_rich"
                    36218..36253
repeat_region      /rpt_family="AT_rich"
                    complement(36734..36768)
repeat_region      /rpt_family="AT_rich"
                    complement(37029..38609)
gene               /gene="At2g15020"
                    /note="T15J14.6; predicted by genscan and genefinder"
                    complement(<37029..>38609)
mRNA

```

Alignment Scores:

```

Pred. No.:      8 96e+03      Length:      114041
Score:          90.50      Matches:         44
Percent Similarity: 39.52%      Conservative: 39
Best Local Similarity: 20.95%      Mismatches: 70
Query Match:      8 4.7%      Indels:      57
                        Gaps:      9

```

US-10-024-955-7 (1-213) x AC005957 (1-114041)

```

QY 28  GULIeAsnLYsAlaIleAspAspAlaIleAlaIleGluInSerGluThrIleAsp 47
      :::::  |||::|  |||:::  :::::  |||
Db 22288  CACTCCAGGCGGCTCTAGATAAAGCTATGTCACAGTCTTAATAGAACGAG----- 22338
QY 48  PromeLYsValProAspHisAlaAspLYsPheGluArgHisValGlyTleValAspPhe 67
      :::::  |||::|  |||:::  :::::  |||
Db 22339  -----AGACGACACTATATTCACATATACAGAGA 22365
QY 68  LYSGLYGLULeAlaMetArgAsnIleGlu-----AlaArgGLYLeuLYSGIn 83
      :::::  |||::|  |||:::  :::::  |||
Db 22366  GAGCTCACCGTTCCTTACCAGATGAGAGAGATTTGGCAACAAGAAAGTAAGAAACCA 22425
QY 84  MetLYsArgInGLYAsnLYsAlaAsn----- 91
      :::::  |||::|  |||:::  :::::  |||
Db 22426  TGGATGAAGAAGAAGATTTGCACACAGAAATTTTTCATGCTTGTACCAAAACAGATT 22485
QY 92  -----ValLYSGLYGLUGLYLeuVal-----LYsAlaHisLeu 103
      :::::  |||::|  |||:::  :::::  |||
Db 22486  TCTGTGAATCCTTTACTGACAAATTAAGATGAGGAGGAATGATTATCGGGGTGACAAA 22545
QY 104  LeuIleGLYValAlaAspAspIleValSerMetGluTYrAspLeuAlaIleTYrLYSGLeuGLY 123
      :::::  |||::|  |||:::  :::::  |||
Db 22546  GAGATAGGGGTTTCAT-----GCTCAGGAGTCTTTACAAAGGTTTATGAGAGC 22593
QY 124  AspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValValAlaLeuSer 143
      :::::  |||::|  |||:::  :::::  |||
Db 22594  AATGTCGTCGCTTTCA-----ATCATGATTTTGGCGATTTCAAACCTCACTAGTACA 22647
QY 144  LeuGluIleSerAspGLUGLYAsnIleThrMetThrSerPheGluValArgInPheAla 163
      :::::  |||::|  |||:::  :::::  |||
Db 22648  GAGCAAAATAATGACGATCTCAAAAGACTTCAGTGAATTTGGAATTT-----TAT 22698

```

```

QY 164  AsnValAlaAsnHisIleGLYLeuSerIleLeuAspProIlePheGLYValLeuSer 183
      |||  ::::  |||
Db 22699  AACACCATTTTCTCATATAGAGATGACAAACACCGGCTCT----- 22740
QY 184  AspValLeuThrAlaIlePheGlnAspThrValArgLYSGluMetThrLYsValLeuAla 203
      |||  |||  |||  |||
Db 22741  GATGCTTAACACGCTCGTTTCTAT-----AAGTCTGTGTGGAGATTGTGCT 22788
QY 204  ProAlaPheLYsArgGLULeUGLYAsn 213
      |||  ::::  |||
Db 22789  CCGGATGCTGAATAAAGAGGTTAAATAAT 22818

RESULT 50
LOCUS      AC098483      177865 bp      DNA      linear      PRI 03-JAN-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-438F14, complete sequence.
ACCESSION  AC098483 AL391534
VERSION    AC098483.2 GI:18042304
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
            Direct Submission
            Unpublished
            2 (bases 1 to 177865)
            Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
            Direct Submission
            Submitted (23-OCT-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
            3 (bases 1 to 177865)
            Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
            Haugen,E.D.
            Direct Submission
            Submitted (03-JAN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
            On Jan 3, 2002 this sequence version replaced gi:16328289.

            ----- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchgs@u.washington.edu
            Drafting Center: SC

            ----- Project Information
            Center project name: chr-1
            Center clone name: RP11-438F14 (sc0161)

            ----- Summary Statistics
            Sequencing vector: plasmid: L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 177789 bases at least Q40
            Consensus quality: 177858 bases at least Q30
            Consensus quality: 177865 bases at least Q20
            Insert size: 177849; sum-of-contigs
            Quality coverage: 7.6x in Q20 bases; sum-of-contigs

            -----
            Overlapping Sequences:
            5': RP11-407H12 (UWGC:sc0688) AL358874
            3': Mapping in progress

```

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdrom@tigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598502.
FEATURES Location/Qualifiers

source 1..114041
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="m1398"
/clone="T15J14"
1..10045
/note="overlap with BAC clone T26120
(AC005396:75170..85214)."
repeat_region complement(60..82)
/rpt_family="(TAA)n"
repeat_region complement(1075..1102)
/rpt_family="AT-rich"
repeat_region 1417..1446
/rpt_family="AT-rich"
repeat_region 1481..1502
/rpt_family="AT-rich"
repeat_region complement(1836..1867)
/rpt_family="AT-rich"
gene complement(3450..7066)
/gene="At2g14970"
/note="T15J14.1; related to En Spm transposon family of maize"
mRNA complement(join(<3450..3728,3810..4172,4473..4769,4847..4885,4927..5217,5452..5532,5790..5993,6080..6673,6899..>7066))
/gene="At2g14970"
CDS complement(join(3450..3728,3810..4172,4473..4769,4847..4885,4927..5217,5452..5532,5790..5993,6080..6673,6899..7066))
/gene="At2g14970"
/codon_start=1
/product="En/Spm-like transposon protein"
/protein_id="AAD03355.1"
/db_xref="GI:4115353"
/translation="MASQAKQVFLREDDSSWHVAMRSPRFRREKENDVTLDIGP
LPTVIDIDVLEKKRGRQPKAAATKCNDEVEYGTLEPTAVBOTVAVENIEVED
IDGETEAEADINGETEAEADINGETEAEADINGETEAEADINGETEAEADIN
AEEPEGELEADINGETEAEADINGETEAEADINGETEAEADINGETEAEADIN
LASVAGALVREHPTITNRMFKIGEIRTLMSVQAKFELEDEOKVAVAKOMGCIN
BSKSRQVTKREARETNOORMLRPNKNVSPFMRKFFVSKTSPEPKIGSADPTVTR
LKWYKSKTKTGTGPKAAETVSSGQSGINTEADSDLSQLGPNPSRLMRGRNMN
KTKLACFQVRSKCMAMQKODQLOQKVNLEQVIDIKIKNHNTCLMLHKAMQSOA
EVGETSAARVNOGSQPKCIIMDMAGTDATVEGCIISDPDEIVNGSRLGPTDVKL
IDTAIVPAYLWRPAIINMEIMKAVGOMIAMPVAMCVSLSEELNPEIDIAOSFVMPAH
KVSVENTDSAVOKAPLOSOSTPTTSASGKSAASOONGATEKSLIGEONCATK
GKSGHNSSTIOVAAAGSESAGTIAKSLGKOPTVYPSHKVKTQLASOTALRSPRQK
TSEGLKANQCKLMDISGKRKRVGAGRHSHIDPDQKHVHRLGENAARWADVAVND
TAVMRPSDEIEMRSLGSSIAMPKDLTVY"
6730..6824
/rpt_family="(CAA)n"
repeat_region complement(7749..10105)
/gene="At2g14980"
/note="T15J14.2; related to En Spm transposon family of maize"
mRNA complement(join(<7749..8219,8355..8663,8888..>10105))
/gene="At2g14980"
CDS complement(join(7749..8219,8355..8663,8888..10105))
/gene="At2g14980"
/codon_start=1
/product="En/Spm-like transposon protein"
/protein_id="AAD03356.1"
/db_xref="GI:4115354"
/translation="MVLGEIARDKKEDDELLAKVDAPLEPLSPGVNHSKLSAIVS
LFRKTNGMSDKSFNDLETLPEMLPEDNVHTSLYEKKKFLSFGMGYOKTHICVN
DCCLERRKYKTLONCLKCNASKTKNTNMTGCVKKYVPOKVLRYTSIVRLKMFSESE
MARDLRMHFNKSTDGKLCHEPDSVYTWDMANKYPLFSEENRNLGLSTNGFNPM
KNTRYSCWPIVLVYNLPPLDLCMKKENITLTLILGSOQPGNSIDVYLEPLIDLCIL

repeat_region complement(13175..13216)
/rpt_family="AT-rich"
repeat_region 13887..13922
/rpt_family="AT-rich"
repeat_region 14418..14533
/rpt_family="AT-rich"
sequence - a consensus. from T23015.64288 to 65702.1415
nt, also present in intron 8 of Arabidopsis thaliana gene
for sulfate transporter GB AB012047"
complement(14535..14596)
/rpt_family="AT-rich"
14726..14753
/rpt_family="AT-rich"
complement(14941..14984)
/rpt_family="AT-rich"
complement(15051..15080)
/rpt_family="AT-rich"
complement(15282..15304)
/rpt_family="AT-rich"
complement(15639..15677)
/rpt_family="AT-rich"
15760..15780
/rpt_family="AT-rich"
16882..16912
/rpt_family="AT-rich"
16934..16962
/rpt_family="AT-rich"
complement(18558..18615)
/rpt_family="(A)n"
18688..18719
/rpt_family="AT-rich"
18849..18895
/rpt_family="AT-rich"
18964..18993
/rpt_family="AT-rich"
19026..19046
/rpt_family="AT-rich"
complement(19873..19908)
/rpt_family="AT-rich"
20381..25506
/gene="At2g14990"
/note="T15J14.3; pseudogene, reverse transcriptase"
/pseudo
complement(20898..20947)
/rpt_family="(CAT)n"
complement(25525..25554)
/rpt_family="AT-rich"
complement(25946..25972)
/rpt_family="AT-rich"
complement(2661..27683)
/gene="At2g15000"
/note="T15J14.4; supported by full length cDNA:
Ceres:19338"
complement(join(<26681..26979,27194..27312,27485..>27683))
/gene="At2g15000"
complement(join(26960..26979,27194..27312,27485..>27639))
/gene="At2g15000"
/codon_start=1
/product="expressed protein"
/protein_id="AAD03357.2"
/db_xref="GI:20197493"
/translation="MAHRNAGSAARSPYSATARSBSLSPTTALPRLRPOSSLPSPR
FTFSSPSRNLALGCTOSFPLPYSVATVSQTLISLNVNLAFCELISNGIDKG"
28132..28169
repeat_region

```

gene
CDS
/translation="MSIKPLHRIYKPIEADENSPGGIVIPDSAKSTKGEVIAG
AGKPLDNGNVRTPCVKGEKVIYGYAGSTYKAEVEYKVLREDDILAIG"
5354..5554
/gene="XF0617"
5354..5554
/translation="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83427.1"
/db_xref="GI:9105489"
/translation="MAECRCFGVGRFCEIHLNPNVQVYCAARVRCADAVGALGK
CKSMOPTLHESLMTGITGYG"
5639..5731
/gene="XF0618"
5639..5731
/gene="XF0618"
/translation="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83428.1"
/db_xref="GI:9105490"
/translation="MVVSGIDALLSAPAVGIDLMKHNGETFMA"
5840..6178
/gene="XF0619"
5840..6178
/gene="XF0619"
/translation="similar to SPI36654 (percent identity: 37 %/query
alignment coverage: 88.4 %/subject alignment coverage:
88.4 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="periplasmic divalent cation tolerance protein"
/protein_id="AAF83429.1"
/db_xref="GI:9105491"
/translation="NASDVIYLFSTCPPLPSEKISRLVQERLACVYOLPGANSTY
RMQGIETQETQILITKTNVAVNAITRLCALHPRLPEALVAVSVGLPEYLTWIN
TEIDVEYSLP"
6175..8451
/gene="XF0620"
6175..8451
/gene="XF0620"
/translation="similar to GI13329039 (percent identity: 32 %/query
alignment coverage: 61.2 %/subject alignment coverage:
67.1 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="c-type cytochrome b1ogenesis protein (copper
tolerance)"
/protein_id="AAF83430.1"
/db_xref="GI:9105492"
/translation="MTPLHRLATLCVLLISPPYVAVSETDLLPYEKAFFVGVKALDR
HHVLYHKIABEYLYIYRRIARVRLNDVLSKDSLNPNGKKKDSITLGDVETIRGS
LTVASVLPVKGVRKLTQVNVQGCADAGVCYPPQVRLHVALPTGNDLDBGAANDR
ODRGSADVAPRLSEIVGVRLPLPADCAFEEAIVGDNHLLRFTPAKGVYAYRD
RTSLAEGDKTVHTGVAWYVPOGSKYHDEHFGVYVFNQIEVMLPLQRDYSGPIANAT

```

Alignment Scores:

```

Pred. No.: 489
Score: 90.50
Percent Similarity: 40.93%
Best Local Similarity: 22.78%
Query Match: 8.47%
DB: 1
Gaps: 12

```

```

US-10-024-955-7 (1-213) x AE003907 (1-10041)

```

```

QY 13 VALAIAValSerAlaSPProIleHIStryAspLysIleThrgluGluIleAsnLysAla 32
|||||
Db 4508 GTGCCCGCTGATGCAATTCATG-----GATCTTAAGCTGGT 4470
QY 33 IleAspAspAla----IleAlaIleIleGluInSerGluThrIleAspPrometLysVal 51
|||||
Db 4469 ATCGATTAAGCAGTATTCGCCGCTTACCAAGCAAGAACATGCTCCAG-----4419
QY 52 ProAspHisAlaSPLysPheGluIleHISValIleValIleAspPheLysGluIleu 71
|||||
Db 4418 CCACACGATGACGATTAACCCATTGCCAAGTCGACACATCTGCTACTCAACGAG 4359
QY 72 AlaMetArgAsnIleGluAlaArgLysGluMetLysArgGlnGlyAspAlaAsn 91
|||||
Db 4358 TCTATCGGCACATTAATTCGACAGCAATGAAGAGTCGTAAGAAGAGATGATTACT 4299
QY 92 ValLysGlyGluGluGlyIleValLysAlaHisLysLysIleGlyValHisAspAsp 111
|||||
Db 4298 ATT-----GAGGAAGGTGACGACTGTAGAAACGAATG-----GATGTCTGC 4257
QY 112 ValSerMetGluTyrAspLeuAlaTyr-----ValIleSerAspIle 120
|||||
Db 4256 GAAGGATGACGATTTGACCGCTGTTACTCTTCCCATATTTCATCAACACGACGATCC 4197
QY 121 -----LysLeuGlyAspLeuHisProThrTrhHis-----ValIleSerAspIle 135
Db 4196 CAATTGTGACTGGATATTCCTTACATCCATCCCTTCAACAGAAATTTCCAGTGTG 4137
QY 136 GlnAspPheValAlaIleuSerLeuGluIleSerAspGluGlyAsnIleThrMetThr 155
|||||
Db 4136 CCGGATTTACTACCGCTGTACCGCGTCGCCAAGAAAGAACGCCGTTGCTGATCTGC 4077
QY 156 SerPheGluValArgGlnPheAla-----AsnValIleAsnHisIleGlyLeu 172
|||||
Db 4076 GCTGAGAGAGTGAAGGCAAGCTTTGGCACTGTGTGTATTAACATCCGCGGATC 4017
QY 173 -----SerIleuAspProIlePheGly-----ValLeuSer 183
Db 4016 ATCAAGTCTGCACTGCAACGACCTGCTGGTATGATGTCGCAACCATGCTGAA 3957
QY 184 AspValLeu-----ThrAlaIlePheGlnAsp-----ThrValArg 195
|||||
Db 3956 GATATGGCTGTGCTGACAGCGGACACGTCGAGTGGAGAAAGTAGTCTGTCTGAA 3897
QY 196 LysGluMetThrLysValLeuAlaProAlaPheLysArgGluLeuLys 212
Db 3896 AAGGCCACGACCAAGCCACTTGCAAGGCCAAGAAAGTAGTCTGTCTTAA 3846

```

RESULT 49
AC005957
LOCUS AC005957
DEFINITION Arabidopsis thaliana chromosome 2 clone T15014 map m1398, complete
sequence.
ACCESSION AC005957
VERSION AC005957.3 GI:20197491
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 114041)
Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Messon,T.M.,
Shen,M., Ronning,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
Unpublished
JOURNAL 2 (bases 1 to 114041)
REFERENCE Lin,X.
AUTHORS Direct Submission
TITLE Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 114041)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission


```

/protein_id="CAGC1475.1"
/db_xref="GI:10639503"
/translacion="MPSDIYHYKQGRFLDYIKDKNYAVAKIDSLVLDLRVAREDDV
LAIPVSDGGLTLIRSAHAILANKVANTLJFDLAPNGVAVENSGFIYDEMKATGEE
DLAIEDEMRKIKENPILREYLAHLYVQKADLLIRFANPKPIRIEDNVDSAPVRSQEN
EVDCLDHPHVSFGYIKAFPLILASAVYKHDESKTIVRIGTAPPEKLIKQYLNMI,
BEARBDHRHRIILAEMLDAVENSEMAAGFELYQYQOTLRKELLASMDMNRKNMFPV,
WTPFPGDTITMKOSGHAKYKRPAMYLFETLPDSDSGIKPMNPGHIAFESRKYSYED,
LPAVSEPTDITVYERKSGEGLTFRATFDGDHAFILMOIGELFTLLTPWKDYAF,
TILGKNEITSLDYSIDRHPREKIVYSYCRNCGYRIGASRTDIECVYCHSHLED,
LISWDNATEQDRLDSMSGIEIKETVGEAEYGPQIDVHAKAIGRMQSTLJDFFE
MPVNFGLTVNSEKEDRVVMIRAIYSEYEVVMAILLEHVAGKLPMLTPIQYVVP
ISSNFEDYARHVAHEILIRHIGISEDITSOETQSKIKLHPYRAYILVSGSKEMENI
SVTAREGSKSYTESFIDVYKKEIEQRKVDAA"
```

Alignment Scores:					
Pred. No.:	2.7e+04	Length:	338100		
Score:	91.50	Matches:	46		
Percent Similarity:	37.38%	Conservative:	34		
Best Local Similarity:	21.50%	Mismatches:	51		
Query Match:	8.57%	Indels:	84		
DB:	1	Gaps:	10		
US-10-024-955-7 (1-213) x TACID2 (1-338100)					
QY	23	Asplysiethrncgluillleasmlysalailaeapspaliaileala-----	38		
	:::	: ::::: ::::: ::::: : ::::: :::::			
Db	76614	GAAMAACCTGTGACAGAGACTGCAGAAGTGCTTGCACATGCCACAGGGCTATGGGTTCAACC	766737		
QY	39	-----AlaIlleGlucIn-----	42		
		:			
Db	76674	CTGCCTGAATCGTTGCTGTGAGAGATAGCTAGAGAACAGACAATGGGGTGAAGAAGGCC	767333		
QY	43	-----SerGIuThrlIleaspromct 49			
		: ::::: :::::			
Db	76734	TACAGAACAATACTCAGACGGTATAGGTGAGGAGATAGCAGAAAGACCTGTAGATCCATPAC	767933		
QY	50	LysValProAspHisAlaAspLysPheGIuAArgHisValGIylIleValAspPheLysGIy 69			
	:::	::::: :::::			
Db	76794	GAAGCC-----GTTGGATCATATAGCGGCTCAGAAC	768233		
QY	70	-----GIuleuAlaMetArGrAsnIIleGIuAlaArgLyLeuLysGI 83			
		: :::::			
Db	76824	ATAAGCGAGACCAAGGTACGACAGATGCGATGAGAGACCTTCCACCTTGGCGGGTTCAGAGGAG	768833		

QY	84	MetlysArg---GInglAspAlaAsnValLysGlyGluGlyIleValLysAlaHis	102
Db	76884	ATGAA-CGTACTCTGGTCTGCCAGGCTCATAGAGTAAGTGA--TGCAGGCGCAT	76939
QY	103	LeuLeuIleGlyValHisAspAspIle-ValSerMetGluTyr-----	116
Db	76940	ACCGAGCACTCCGTCATAGCAGTCTACCTGAGGCCGACGATACGACAAATGACGAAGT	76999
QY	117	-----AspLeuAlaTyrLysLysGlyAspLeuHisProThrThrHisValIleSerAs	134
Db	77000	CGTGTGATGTGTGTGGAAGAGCTAGAAATACAAAGC-----ATAGACGA	77044
QY	134	pIleGlnAspPheValAlaLalaLeuSerLeuGluIleSerAspGluGlyAsnIlePheMe	154
Db	77045	TGTTGCAGATATA-----ATACTGATATAGGGGAGCTTACCGT	77083
QY	154	TThr-----SerPheGluValArgGlnPheAl	163
Db	77084	TACCGTGAAGGCCAGATCCGAGAAGACAAAGACAGCTAATTCAGAGATGAAAGACATAA	77143
QY	163	aAsnValValAsnHisIleGlyGlyLeuSerIleLeuasp	176
Db	77144	GAAACGCAATTAAGCAAGATTAAGGCGCATCCGTAATGAG	77183
RESULT 44			
AF436987			
LOCUS	AF436987	558 bp	DNA linear BCT 09-JUN-2002
DEFINITION	uncultured pig faeces bacterium clone 004_H12 60 kDa chaperonin (cpn60) gene, partial cds.		
ACCESSION	AF436987		
VERSION	AF436987.1	GI:21359045	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Hill,J.E., Seip,R.P., Betts,M., Hawkins,L., Van Kessel,A.G., Crosby,W.L. and Hemmingsen,S.M. Extensive profiling of a complex microbial community by high-throughput sequencing Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)		
TITLE			
JOURNAL			
MEDLINE	22034982		
PUBMED	12039767		
REFERENCE	2 (bases 1 to 558)		
AUTHORS	Hill,J.E., Seip,R.P., Betts,M., Hawkins,L., Van Kessel,A.G., Crosby,W.L. and Hemmingsen,S.M. Direct Submission		
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			
BASE COUNT	135 a	101 c	179 g
ORIGIN			

TITLE The genome sequence of the thermophilic scavenger Thermoplasma acidophilum
JOURNAL Nature 407 (6803), 508-513 (2000)
MEDLINE 20479972
PUBMED 11029001
REFERENCE 2 (bases 1 to 338100)
AUTHORS Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. and Baumeister, W.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-2000) Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG Bioinformatics, Smith Kline Beecham Pharmaceuticals, Collegeville, Pennsylvania 19426, USA GSF-Forschungszentrum fuer Umwelt und Gesundheit, Munich Information Centre for Protein Sequences (MIPS) am Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG Project Coordination: Andreas Ruepp and Wolfgang Baumeister E-mail: ruepp@biochem.mpg.de, baumeister@biochem.mpg.de
COMMENT Information on performance of analysis and a more detailed annotation can be viewed at:
http://www.biochem.mpg.de/baumeister/genome/ and at:
http://pedant.mips.biochem.mpg.de/.
FEATURES
Source Location/Qualifiers
1..338100
/organism="Thermoplasma acidophilum"
/strain="DSM1728"
/db_xref="taxon:2303"
115..1704
/gene="Ta0324"
115..1704
/gene="Ta0324"
/note="similarity to known protein: O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [validated] - Pasteurella haemolytica (serotype A1); PIR2:A38108"
/codon_start=-1
/transl_table=-1
/product="O-sialoglycoprotein endopeptidase related protein"
/protein_id="CAC11469.1"
/db_xref="GI:10639497"
/translation="MIVGLGMAHRTISGHIIDESIIAMESSMYRPKGTIRPLDAA VHSVIDIVIRALEKAKISIHDDILISFGSGPLAPSLRTATARTISVITCKPI IGVNPLHIEIGRRTGAIDVPMILVSGNTVIAHVNGRVRLGETLIDIGINMID KPAKEGIPFGPGEIEKTLAMKGTLLDLPYVKGMDTAFSGILTAALYKTAQAE DISYIOETAFMLVEERLALVSGKDEILMAGVALNRRLDMVTNNARAGISY LDRREYCMONGIMIAQALIMKSGVMSVEEAVNPRFIDEVADPWTIDSRKYG KAGASRIEVEFPHGRPARIRKVRISKSYNSPLDKIRYERRMNETLIRKIKGAGN SPYVYDFDFPFSSTWQIKIPGNMSEELNEGRTDFINELGIMAKHRRAGIAHGLTV NNIYNDVFIIDPSMGKYNAEIEDVADIVALEDSIKGLDGSVIGQMLKSYRNN FNIADVLETVSAIRRRHYV"
1697..2260
/gene="Ta0325"
1697..2260
/gene="Ta0325"
/note="similarity to known protein: 6-N-hydroxylaminopurine sensitivity-controlling protein HAM1 [validated] - yeast (Saccharomyces cerevisiae); PIR2:S57088"
/codon_start=-1
/transl_table=-1
/product="HAM1 protein related"
/protein_id="CAC11470.1"
/db_xref="GI:10639498"
/translation="MYRVTNRRHRRPEEVSMAALGIEIEVMRMKYEELDDSTPERI SYDSCNKSRIYEAPEYDSCGLFIALRGPGYSPNYVSTPTGNGCIIKIMDDNDR TAYFLTVSMNGHSTIOTPGYMGKIARSINGSGNGFGIDPIFIPGSDRTAEMDK AKNAISHRSIAFRGLFEYIKMOTAK"

gene
CDS
/transl_table=11
/product="fixC protein related"
/protein_id="CAC11471.1"
/db_xref="GI:10639499"
/translation="MSEFVIVTIGAGPACSSAAIKIAGCGNVLLVERGPPGTRKNS GAVLMDELSKVPDMERSAPRIERYTRKEVAFLEKSKIAISYSQELMEKSKITI VVGKTDQMLARKAKDAGALVGTIVDKYIADIGKVIYGMQGDVTVDAVVAEAGN PVLVINSGLRPKLSHDVVALGKETIKLEPNYINERFNDSSGGAASEPVVLELDGL LSGGFLYTRDFTISLGIYVSLDRLRANGMTHSYDINGKFEDEHPFIKIDGIPDEYA AHLVPEGGLASMPKLYGNGYLVAGDAAMPTFSNGVYLOGINATISGMLAEITLENK GKISEFTLSTYYSKLRNSVLSLDTTFNCTQVTVSDFQKTMPIKIDSVMSMDEER GMPKRLMOVLTAAVRSNGMKTDDLISGTRMRKX"
3747..4019
/gene="Ta0327"
3747..4019
/gene="Ta0327"
/note="strong similarity to unknown protein: ferredoxin - Thermotoga maritima (strain MSB); PIR:A72242"
/codon_start=-1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC11472.1"
/db_xref="GI:10639500"
/translation="MKIEKLSLANKYTRDSYAHITIDPDICAGCPDHFVCPACPANC YTLINKLNFKYEDCVCEGCTDIACSHGSVKWTLPGKNGVIYKYG"
4044..4820
/gene="Ta0328"
4044..4820
/gene="Ta0328"
/note="similarity to known protein: FIXA PROTEIN, FIXA PROTEIN - Escherichia coli, Escherichia coli; SWISSPROT:FIXA_ECOLI"
/codon_start=-1
/transl_table=11
/product="FIXA (related to carnitine metabolism) related protein"
/protein_id="CAC11473.1"
/db_xref="GI:10639501"
/translation="MIKQVPSSEVEIDPVKMTLNTRKARNVNAADLNALEALRIK DVGVAIVTISMGPPMADAIIECMARGDRCGLITDRAPAGADYTPGTLTAATTK LGNDVIFRGDETTOSTGTHVCGPAERLGGPOITAKKVDYEDGVYITRPLEDDK IYKVPPIVATVLLNSNIPRHOSLRKIDAIKRGVSKCINELGLPEWVGIRGSPFI VSMKAIRKEPEHLHKTVTLDKIDIVADLVSKNIIKVGK"
4817..5872
/gene="Ta0329"
4817..5872
/gene="Ta0329"
/note="strong similarity to known protein: ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNIT (ALPHA-ETF) (ELECTRONTRANSFER FLAVOPROTEIN LARGE SUBUNIT) (ETFLS) - Clostridium acetobutylicum; SWISSPROT:ETFA_CLOAB"
/codon_start=-1
/transl_table=11
/product="probable electron transfer flavoprotein, alpha subunit"
/protein_id="CAC11474.1"
/db_xref="GI:10639502"
/translation="MMVGLINALIPVANKRDEKYNQVYIEHRDEIKRSLEMGIGCKO LARKINEKLCVYIGDKVDGIAKEVGEYCDIVLAESPDIAEIRMPAQLIGDVIYA EYKPNIFLTAATRNGDNLASRIAVERETITADCTVLDENKTLIANPPTYGESTLA EILCRNHRPOMATNARGPTTPPAKEKDHRYEIVKRVNDVDMKMLKKQIIFKPIPKATD LTAAKIYVAGGIGLGGPDGFKLDDIADLIGAVGSSRPVLDGWTTRDHOVGOTGOA VRPDLVAVAGISGKPDHVAQMFYSKIISINQDPAEINKISDYIIVEDYRAAVPALI DAIKNGKQKVAEAKA"
complete(5860..7842)
/gene="Ta0330"
complete(5860..7842)
/gene="Ta0330"
/note="similarity to known protein: threonine--tRNA ligase (EC 6.1.1.3) - human; PIRI:YSHU9"
/codon_start=-1
/transl_table=11
/product="threonine--tRNA ligase related protein"

Oy 168 -----HisIleGlyIy-----LeuSerIleIeu 175
Db 9987 CSTACGAGAGACGGTGTGAAGCCAGCTGTTCCGCTTTATTTGGGAATTACG 10046
Oy 176 AsproIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArg 195
Db 10047 AAGCTTCTTTGGGTACGGAATCGTTATATCATCAGACGCTTCTTCCAGACACA----- 10100
Oy 196 LysGluMetThrLysValLeuAlaProIa 205
Db 10101 -----ACTCGACTTTTACACATGCA 10121
RESULT 42
LOCUS AP002545
DEFINITION Chlamydia pneumoniae J138 genomic DNA, complete sequence,
ACCESSION AP002545 299650 bp DNA linear BCT 25-MAY-2002
KEYWORDS Chlamydia pneumoniae J138
SOURCE Chlamydia pneumoniae J138
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1
AUTHORS Shirai, M., Hirakawa, H., Ouchi, K., Tabuchi, M., Kishi, F., Kimoto, M.,
Takeuchi, A., Nishida, J., Shibata, K., Fujiwara, R., Yoneda, H.,
Matsushima, H., Tanaka, C., Furukawa, S., Miura, K., Nakazawa, A.,
Ishii, K., Shiba, T., Hattori, M., Kuhara, S., and Nakazawa, T.
Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and
the United States
J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)
20298966
JOURNAL MEDLINE
REFERENCE 2
AUTHORS Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K.,
Shiba, T., Ishii, K., Hattori, M., Kuhara, S., and Nakazawa, T.
Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and ChL029 from USA
Nucleic Acids Res. 28 (12), 2311-2314 (2000)
20330349
JOURNAL MEDLINE
REFERENCE 3
AUTHORS Shirai, M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
School of Medicine, Department of Microbiology, 1-1-1
Minamikogushi, Ube, Yamaguchi 755-8505, Japan
(E-mail:shirai@epo.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
Fax:81-836-22-2415)
COMMENT On or before Aug 31, 2000 this sequence version replaced
g1:6172286, g1:6172288, g1:6172310, g1:6172312, g1:6172314,
g1:6172316, g1:6172318, g1:6172320, g1:6172322, g1:6172324,
g1:6635158, g1:6635160, g1:6635162, g1:6635164, g1:6635166,
g1:6635168, g1:6635170, g1:6635172, g1:8547422, g1:8547426,
g1:8547430, g1:8978373
FEATURES
Source location/Qualifiers
1. 299650
/organism="Chlamydia pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
complement(1..282)
/gene="CpJ0001" .282)
complement(1..282)
/gene="CpJ0001" .282)
complement(1..282)
/transl_table=1
/codon_start=1
/product="CT001 hypothetical protein"
/protein_id="BA098211.1"
/db_xref="GI:8978374"

/translation="MLKIIINGLSLIIICALNVLIGITHNNKINIITAKLCGGVSTP
ATQIYIIIGIAGVICLLSFCPFESKSRSHSGDSCSSGGCHSHSDKN"
573..878
/gene="gac"
573..878
/gene="gac"
/codon_start=1
/transl_table=1
/product="Glu-tRNA Gln amidotransferase (C subunit)"
/protein_id="BA098212.1"
/db_xref="GI:8978375"
/translation="MEQPHLDREIILLAKASALDSEELIOETSLSAVTSKKA
LAIEDDADSCESLPMHVNVEDLREDSVTSDFNREFLRNVPESLGLVYPVAVIK
895..2373
/gene="gata"
895..2373
/gene="gata"
/codon_start=1
/transl_table=1
/product="Glu-tRNA Gln amidotransferase (A subunit)"
/protein_id="BA098213.1"
/db_xref="GI:8978376"
/translation="MYRSALSLAKAVTIGELTANGVONHPPHREAEAGVAPISL
CKEALDEALIDKRRSGEPGLAGVPVGIKDNIVHTGLKTTCAASVLEVOGPPD
ATVERIKEDGIIIGKLNMDPEFAGSTFLVSAFPTNPMDLKSVPGSSGSAAY
SARCPVALGSDPTGSGIRPAPAFGCVGFKSPGAVSRVGVAFSSLDQIGPLANT
EDVALMDVFGSDPKDATSRFEFDSFMSKISTGPVGVIGVPTFLGLDIDIENT
FSLIAFEGEGHLVDLILSHAVSYIYLIASEAATNLARFGVYGGVRRPAHT
ISQIYDLSRGSGFGKVEWMRLILGNVYSARONYYKKAIVARKIKYKARPTAEC
EILAMPVSSPAFELIDPVTTLTLDIYVAMLAIVPALAVPSGSKGLPLGLQ
IIGQGQDQOVQCVGVSFOEHAQIKQLFSKRYASVVLGGOS"
2370..3836
/gene="gatb"
2370..3836
/gene="gatb"
/codon_start=1
/transl_table=1
/product="(Pct112) Glu-tRNA Gln amidotransferase (B
subunit)"
/protein_id="BA098214.1"
/db_xref="GI:8978377"
/translation="MSAYADMESYIGLEHYVELNTASKLFSSALNRGDEPNINIST
VCTGLPGLPVLNDSVEREAVLFGCAVGEISLSLRFPRKSYPPDSRNFOIOTFEPH
PIINGRKAIIVOGERYEFLAOTHIEDDAGMLKHFGEAGDVNRACVPLIEIVSKP
CMCEPEDAVAVATSLVSLDIYIGSDCMERGSIFEDVNAVSRPGSPELNRIKESAE
MNSFAPMAALAEAKOROIDEYLNPNKIDPLVIPAIVYRMDPEKKIVLNRIKESAE
DYKTRPEPDLPTLQTESYTERIKTIDELPYDKHKRIQERYGSESDIASLISKNI
ATFEVACKDKNFRSLSNWVSVERGGCKTLGVLPSSGIFPEGVAVLVAIDGVI
TGKIAKEIADLMESPGKNPEILKEPELLPMSDEGLOKITAEVLANPESIVDYK
NGKTRALGFLVGOIWKRTAGKAPKRVNELLLILDG"
4127..6895
/gene="pmp.1"
4127..6895
/gene="pmp.1"
/codon_start=1
/transl_table=1
/product="polymorphic outer membrane protein C family"
/protein_id="BA098215.1"
/db_xref="GI:8978378"
/translation="MRSLSGCPPLVSEFTLLSVFDTSLSATTTISCTPDSRHGSQNA
ERSYVNOAGDVSILGDAVISNVDSALNKACFNTVSGSVTFAGNHGLYNNISSTGT
TKEGVILCCDPOATARFSGSTLSFIOSPDIDKEOGLYSKMLALNNVYREFEON
OSKTRGKAIISGANTVIVGNDVSFYOAAATFGGAIHSSGPLQIIVNOAETIRFQONTA
KNGSGGALYSGDIDIDQNAVVLPRENLATTAIGACVCCLOPLPSSGSTVPIYTES
DNKQVLEPRNHSIMGGAIYARKLSISSGGPLTFNNISYANSONLGAIALIDGGEI
SLSAEKGTITFOGNTSLPLTNGIHLONAFPLKIQANNGSISIEYDPTISBADSSTO
LNINGDPRNKEVYTGIIIFSGERSLANDRDKSTIPQNVNISAGLVITKEAEVYVSK
FTOSPGSHLVIDLGTKLASKEDIAITGLADISLSSSTPAAYIKANTAKQISVTD
SIELISPTGNAYEDLMRNSOTFPLLSLEPAGSGVITAGDVLDPVSPHYEQGMKL
AMTGTGNKVGFEFMDKINYPKPEKEGLVNIILGMNAVDRSLDVOETHASSILOTO
RGIAIDGIGNEFHVYASRDNIRYHNSGGLVSVNNETTPKRPMSANLSQLESDKQY
AVSNNEYPMYIGSLIYQTTISLGNITFRASRNPNVNGITLSRRPLMLKTFPHLCAY
GHATNMDKTYANFPMVKNVNRNMCWALTECGSGPLLVFENGRLFOGALIPMKLDLVY

```
gene
1250..1762
/gene="r110"
/ncbi="CP0079"
1250..1762
/gene="r110"
/codon_start=1
/transl_table=1
/product="L10 Ribosomal Protein"
/protein_id="AAD18232.1"
/db_xref="GI:4376337"
/translation="MKQKTLTLOEVEKISAGCFILRLRLPTAASREPRNSLSG
VSAEVEFKRIFFKALAEAEVDCSDTDHGLGVFSCGDPVSAKQVLFNQHDKD
SLVLAGRMNDAASGAEVAVAKPLSEKLRLQOVVGLFAAPMSQVAVGIMNSVGLI
SCVQKAGKN"
1806..2195
/gene="r17"
/ncbi="CP0080"
1806..2195
/gene="r17"
/codon_start=1
/transl_table=1
/product="L7/L12 Ribosomal Protein"
/protein_id="AAD18233.1"
/db_xref="GI:4376338"
/translation="MTTESLETVEKLSNLTVLIELSQLKLEEKMDVTASAPVAVA
AGGGEAPVAAPTEFAVTLIEDVPADKKIGVLKVVREVTGLAKKAKEMTGELPKTVK
EKTSKSDAEDPVKKLQDAKAKAFKGL"
2332..6090
/gene="r208"
/ncbi="CP0081"
2332..6090
/gene="r208"
/codon_start=1
/transl_table=1
/product="RNA Polymerase Beta"
/protein_id="AAD18234.1"
/db_xref="GI:4376339"
/translation="MLKPERVSVAKKEDIPDLNLIETQIKSYKQFLQIKLAERE
NIGLEEVREIFPIKSYNEATVLEIYSLNPKVPSPECIRGIVTVTLKVRRLT
DEGTGKEEVVMTGPIPLMTDKGTPLINCAEVVVVSOVIRPCINPEOKHSKNLTES
FRITPYRGSMLAIFDINDLIYIHIDRKKRRKRIATITFALGTSSDADIIEEFTLI
GESSLRSEKDFALLVGRILADNIIDEASSLYGAKGELSTAMLRMLADIAEYKIA
VADENHPIIKMLAKDPDYSALDKEDYRLRDEPATTANARSTIKRLFEDPKRYN
LGRVGRYHLKRLKELRSNIDDEALSOVTLKEDVIGALKYILRLKMGDEKACVDDIHLA
NRVRSVGEELIONOCRSGLARMEKIVREKMLFSDSLTPGKVVSAKGLASVLCIET
FGRSLSQPMODTNPVAILTHKRRSLAPGCLNBERGCFEVRVNAASHVGRICET
PEGNIGLITLSLSPAKINEGFIETPRIVRDCIVDEIEYMRADVEEVEVIAOASA
SLDERNMTEPCWVRVAGEAFEDATSTVTHMDSVPKQVSIIVGLIFELPHDANRA
LMGSNMQOAVPLKTEAPVVGTELECAADSGAIVAEEDGVVDVFDGKVVVAAK
HNPTIKRTYHLKELKELRSNIDDEALSOVTLKEDVIGALKYILRLKMGDEKACVDDIHLA
LVAPMPWGVNFEADAIIIEKRLINEDATSYIEEELTADTCLGKEEITRDLPKNS
DEVILANGEDGILIGAEVKKPGDILVNGKITPKSETELAPEKRLAITEGKAADYKDA
SLVYPTGEGVMDVYKVSRRDKLSKSDDELVEEVAHNLKDKQKKNQVATLKYERE
KUGALLLEKAPPAIIRKTAIVHESLTDQETIERIEBEDVLDMPCEMEVL
KGLSDYETALORLEINIKTEVENHIREGDADLDHGVIRQVAVYASKRLQVGDKMG
RHGKGVYSKIVPEADMPYLSNGETVOMILNPLPGVPSMNLGOVLETHLGAAKTAGI
VYKTPVEGFEPEORLMDMILBOGLEPDKSPLDGTGERDNKVYIYIYIMLKLSHL
IADKIHARSIGPSYLVTOQPLGKRAOMGGORGGEMEVALLAECYAHNLQELLTYKSD
DVSGRTRIVESTIVGENILRSSTPESFNVLIKEMOGLGLDVRPMVYDA"
6114..10295
/gene="r208"
/ncbi="CP0082"
6114..10295
/gene="r208"
/codon_start=1
/transl_table=1
/product="RNA Polymerase Beta"
/protein_id="AAD18235.1"
/db_xref="GI:4376340"
/translation="MGENSRDGLVLSKEGLFDKLEIETIADITIRDMWSGELIKPE
TINRTRPEKGLFCEKIFGPTKDWECCKGKTKIKKIGVLCDCGCEVLUSKRRR
RMAHLEAVPIVHIFEFKTRISGIVNIGMASDLEERYIYEYVVIDPGKTDLLKQO
LLNDAOYREVVEKMGDAFVAKMGSEAIYDILLKSEDLQSLKLKLEKRLKTKSQOARM
```

```
KLARKLKIIIEGVSSSNHPHEMVLKNIIPVBPDLRLVPLDGRFATSLNDLYRRLI
NRNNRLKAILRKTPEVIVRNEKRLMDQAVLADLPNGHHRHAGNRPKLSISEML
KGNKRFQNLILGRVYDSCRSVITVGGELFENOCGLKEKALELEPEFLIKRLDOG
SVYTRSAKKMIIORGAPRPMVDLEIILKGHVLLNRPATLRLHQAEPVILIEGKAI
RIHPLVCAFAFNADDEGDMAYHVPISVAQOIEAKVLLMAMPNITLPPSSGKVALPSKD
MTGLUYILMADPTYPEEHGGKTKIFKDEIVLRLANNGGFIIDVFPGRDGTGRGIR
IHERIKVRIIDQIIEETPGFVRLNRYVPEKELFQNSMPSKRISELILQCYKQVLEA
TVRFLLDLDKDGFIQATKAAISMGLKQVRIEDIKSHLIDKADYDVAIVKQYDGIIT
EGERNKTSISIMTEVSBOLSDALVVEISKORSNPLPFLINDSARKNKLSQLOLA
LRGIMAKNNGALIESPITSNREGIVLEYSISSHGAKGLADIRKADSGYITRLRL
LVADQDVITREKDGTLNHIETSAIGQSEBLLPKDRITGRVAEDYTOFGDYSRLI
AOSGDVNSVQAEALIDAGIETIKIRSTLGEISPGVCACQGLNLAGRLIGGEAV
GIIAOSLIGEBGTQIIMRTFHLGIIAATSTPEIITNSGIIIVMDLIVGOGCHNL
VLNKKGLAHVVGDEGRITLNEKTKLSTKSIESTLEFVPELVGLIIVADGTPVSGORI
GVELNHPITICDRGFTIKYEDIVGISTEELVYVKNNGTVELIYKOHNGELHPQIATY
DDADSELVGTAYALPSGALISVEGQARDPMLLARLPKRAIKTKDITGGIPRAVELY
EARKPEDADIAKIDGVNDFKGIQNKRIILVOCBMKEEHEHLIPLKHLIVQDGS
VIRGOQLDGLVNPHEILLETIGVRELQVLVNEQVEYRLOGVRLQGVNKHIEIIVROML
OKVRTDPGDTTLTFGEDVKNRKEFEYENRRFEEDGQPAQAVPVLGITTKASLETSEF
ISAASFODTTRVLINDAACCSKTDVILGFKENYINGHMPGCTGFETHKRIQYLEKEQ
EDLVPDFSETECVYC"
BASE COUNT      3263 a      1837 c      2288 g      2990 t
ORIGIN
Alignment Scores:
Pred. No.:      417      Length:      10378
Score:          91.50      Matches:      53
Percent Similarity: 35.20%      Conservative: 35
Best Local Similarity: 21.20%      Mismatches: 63
Query Match:      8.57%      Indels:      99
DB:              1      Gaps:      12
US-10-024-955-7 (1-213) x AE001593 (1-10378)
QY      36 AAlaIleAlaIleGluGlnSerGluThrIleAspPro----- 48
Db      9429 GCGATTATCTCTGTAGAAAGAACAGCAACGCGTTGATCCAGTATGTTAGCTAGACTT 9488
QY      48 ----- 48
Db      9489 CCTCGCGGAGCTATCAAAACAAAGATATTACTGGCGGTTTCCCTGCTGCTGAATTA 9548
QY      49 -----MetIysValIProAspPHisAlaAspLysPheGluArgHisValGlyIleValAsp 66
Db      9549 GTAGAAGCGTCGTAACCGTAAGCAAGATGCTGCTGACATCGCCAAATGATGATGCTGTGTGAC 9608
QY      67 PheLysGlyGluLeuAlaMetArgAsnIle-----GluAlaArgGlyLeu 81
Db      9609 TTCAAAAGGAATTCAAAAGAACAAAGCATTTCTGTGTCTGTGATGAATGACAGGTATG 9668
QY      82 LysGln-----MetLysArgGlnGlyAspAlaAsn 91
Db      9669 GAAGAAAGAACATCTGATTCATTACCAACAAACTTTGATGTTGACAGGATGAGATGATG 9728
QY      92 ValLysGlyGlu-----GluGlyIleValLysAlaHisLeuIle----- 105
Db      9729 ATTTAAAGGACAGACAGTTCACGATGTTAGTGTTCCTCATGAATCCTAGAAATTTGCG 9788
QY      106 GlyValHisAsp-----AspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGly 123
Db      9789 GAGATTCTGGAACCTTCAGAGTACCTGTAATGAGAGTGCAGAAAGTTGACGCTGCGAG 9848
QY      124 AspLeuHisProThrThrIleValIleSerAspIleGlnAspPheValValAlaLeu---- 142
Db      9849 GCGGTT-----GACATTACGATGATGAAGCATTTGAAATTAATT 9884
QY      143 -----SerLeuGlnIleSerAspGluGlyAsnIleThrMetThrSer 156
Db      9885 GTTCGTCACATGTTACAAAGATGACGAATTTACTGACCCGAGTGAATGACGCTGCGC--- 9941
QY      157 PheGluValArgGlnPheAlaAsnValAsn----- 167
Db      9942 -----TTTGGCAAGACGTCGATTAAGAAAGACTTTTATGAAAGAAATCGT 9986
```

```

similarity: putative"
/codon_start=1
/transl_table=1
/product="ribosomal protein L1"
/protein_id="AAF38505.1"
/db_xref="GI:7189611"
/translation="MTKHKRIRIGILKNYDESKSYSLREADIIKOCPPVRFDTVDV
SIKIGIDPKSKDOOIRGAVLPNGTKTLRLIVYASGKNKVEAYEAGADPFGSDLYE
KIKSGMLEFDVAATPDMREVGKLVGGRNRLMPPTKGTGTTVDVAKAISLRKQ
IEFKADRGVCNVGKLSFESSQIKENIEALSALIKAPPAKGGYLVSFITISSTM
GPGISIDIREIMAS"
complement(9873. .10298)
/gene="Cp0698"
/complement(9873. .10298)
/gene="Cp0698"
/note="similar to GB:DJ3303 SP:006796 PID:285629
GB:AL009126. identified by sequence similarity; putative"
GB:AL009126.1
/codon_start=1
/transl_table=1
/product="ribosomal protein L11"
/protein_id="AAF38506.1"
/db_xref="GI:7189612"
/translation="MSYKVKIKIIRKQIPGSKANPAPPICGALCAAGVNIQFCKEEN
AATODKPEDLLPVVITVADKTFPTTKOPVSSLIKKTLMESGSKIPNNKKVGLT
QAQVEALAEQMKMDIVLESARMEVGTARSMGIDVE"
BASE COUNT      2989 a      2287 c      1837 g      3263 t
ORIGIN

Alignment Scores:
Pred. No.:      417      Length:      10376
Score:          91.50      Matches:      53
Percent Similarity: 35.20%      Conservative: 35
Best Local Similarity: 21.20%      Mismatches: 63
Query Match:      8.57%      Indels:      99
DB:                1      Gaps:      12

US-10-024-955-7 (1-2113) x AE002228 (1-10376)
Oy 36 AAlaIeAlaIaIleGlucInSerGluThrIleAspPro----- 48
    ||||| :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 953 GCGATTATCTCTGTAGAACAGACACACGGGTTGATCCAGTATGTTGACTAGACTT 894
Oy 48 ----- 48
Db 893 CCTCGCGAGCTATCAAAACAAAGATATTACTGGCGGTTTCCTCGTGTGCTGATTA 834
Oy 49 -----MetLysValProAspHisAlaAspLysPheGluArgHisValGlyIleValAsp 66
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 833 GTAGAGCTCGTAAACCTGAAAGATGCTGCTGACATCGCCAAATGATGCTGTTTGAC 774
Oy 67 PheLysGlyGluLeuAlaMetArgAsnIle-----GluAlaArgGlyLeu 81
    ||||| :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 773 TTCAAGGAATTCAAAGACAAACAGCTATTCTTGTGCTGTGATGAATGACAGGTATG 714
Oy 82 LysGln-----MetLysArgGlnGlyAspAlaAsn 91
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 713 GAAGAAGAACATGATTCATTAACCAAACTTGTGATGACACGTCGACATAGTGTG 654
Oy 92 ValLysGlyGlu-----GluGlyIleValLysAlaHisLeuIle----- 105
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 653 ATTTAAAGGCGACAGCTTACCGATGTTAGTTGTTCCCTAGAAATCTTAGAAATTTGC 594
Oy 106 GlyValHisAsp-----AspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGly 123
    ||||| :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 593 GGAGTTCTGGAACCTCAGAGTACCTGGTAAATGAGGTGACGAAGTTTACCCTGACG 534
Oy 124 AspLeuHisProThrThrHisValIleSerAspIleGlnAspPheValAlaLeu--- 142
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 533 GCGGTT-----GACATTAAAGCATTAAGACATATTCGAATTTATTT 498
Oy 143 -----SerLeuGluIleSerAspGlyGlyAsnIleThrMetThrSer 156
    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 497 GTTCGTGAGATGTACAAAAGTAGCAATTAAGTACCGCAGGTGATACGACTCTGCTC--- 441

```

```

Oy 157 PheGluValArgGlnPheAlaAsnValAlaAsn----- 167
    ||||| :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 440 -----TTTGGCCAGACGCTGATTAAGAAAGACTTTTATGACAAATCGT 396
Oy 168 -----HisIleGlyLys-----LeuSerIleLeu 175
    ||||| :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 395 CCGTACGAAGAACAGCGTGTAAAGCAGCTCAAGCTGTCCCGTCTTATTTGGAAATTACG 336
Oy 176 AspProIlePheGlyValLeuSerAspValLeuThrAlaIlePheGlnAspThrValArg 195
    ||||| :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 335 AAGCTTCTTTGGTACGGAATCGTTTATATGACGACGCTTCTTCCAAAGACACA----- 282
Oy 196 LysGluMetThrLysValLeuAlaProAla 205
    ||||| :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 281 -----ACTGAGCTCTTAACAGATGCA 261

RESULT 41
AE001593      10378 bp      DNA      linear      BCT 01-DEC-2000
LOCUS
DEFINITION Chlamydia pneumoniae section 9 of 103 of the complete genome.
ACCESSION AE001593 AE001363
VERSION AE001593.1 GI:4376334
KEYWORDS
SOURCE
ORGANISM Chlamydia pneumoniae CWL029.
REFERENCE 1 (bases 1 to 10378)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
COMPARATIVE GENOMES OF CHLAMYDIA PNEUMONIAE AND C. TRACHOMATIS
Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PubMed 10192388
REFERENCE 2 (bases 1 to 10378)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1..10378
/organism="Chlamydia pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
84..509
/gene="r111"
/note="Cp0077"
84..509
/gene="r111"
/codon_start=1
/transl_table=1
/product="L11 Ribosomal Protein"
/protein_id="AADI8230.1"
/db_xref="GI:4376335"
/translation="MSYKVKIKIIRKQIPGSKANPAPPICGALCAAGVNIQFCKEEN
AATODKPEDLLPVVITVADKTFPTTKOPVSSLIKKTLMESGSKIPNNKKVGLT
QAQVEALAEQMKMDIVLESARMEVGTARSMGIDVE"
536..1234
/gene="r11"
/note="Cp0078"
536..1234
/gene="r11"
/codon_start=1
/transl_table=1
/product="L1 Ribosomal Protein"
/protein_id="AADI8231.1"
/db_xref="GI:4376336"
/translation="MTKHKRIRIGILKNYDESKSYSLREADIIKOCPPVRFDTVDV
SIKIGIDPKSKDOOIRGAVLPNGTKTLRLIVYASGKNKVEAYEAGADPFGSDLYE
KIKSGMLEFDVAATPDMREVGKLVGGRNRLMPPTKGTGTTVDVAKAISLRKQ
IEFKADRGVCNVGKLSFESSQIKENIEALSALIKAPPAKGGYLVSFITISSTM
GPGISIDIREIMAS"

```

[illegible]

AELVBARKEPPEADIAKIDGVDPFKGIQKNNKILVYCOBMTGMEHEHLIPUTKLILVO
 RQDVIYKGOQLTDLVPHHILKEICGERELQYEVLRNVEQEVYRLOGVDINKRLIEIV
 ROMLOKAVITDPGGTITLIFEDYKKEFEEENRKEPEDEGGCPAQAQVPLVLTGITSAGJ
 TESFISAAFOEDTTRVLTDAACSKSDYLLGKENVIMGHIMPGTGSETKRIKOYL
 EKEDEIVLAFDVTSTEEVCC^a
 complement(4292. .8080)
 /gene="CP0694"
 complement(4292. .8080)
 /gene="CP0694"
 /note="similar to GB:A6001273; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="DNA-directed RNA polymerase, beta subunit"
 /protein_id="AAF38502.1"
 /db_xref="GI:1789608"
 /translation="MMSNQRPRMLKCEPVSUKKEDIIDPLMLTIOIKSKOFL
 OIKLAERENIGLEEVFRITPIKSYNEATYELTISRLGVYKSPSECTRREITTS
 VTLVRFPLDDETGICEEYVMTPIYMDKQTFIINAEERVYSVHRSGINFEDE
 KHSKNILFESRIIPVSGMSEIIPINDLIYIHDIKRRKKILATIFIALGSSD
 ADIIIEEFTTIGESLRKEKFOALLVRIIADNIDBASLVYKAGEKLTSMALRML
 DAGIASIAVADADENHPIIKMLKAPPTQSYEALKDFPYRLRPEEPVLTANASTJ
 RLFDPRYRNGVRGRKILRKIGESIDDEALSOYTLKREYVATLPIKMGDKK
 ACVDIIDLANKRVRYSVGEILINQCSGGLARKEKLYREARLMPDSSTLTPGKVSAA
 KGLASVLDFEGRKQLSQFMDQNPNAVELTHRRISALGPGGLNERRGFEFVRVHAS
 HGRYCRLEPEDEGNILITLSSEFAKINFEIGITPYAVNDGLVTDIEIVMDADVE
 BEYVIAQASAEIDEDYNNFTSPVCSVYAGEAFADSTVYTHMDVSPKQLSVITGLPI
 FLHEDDARALMGKQMOARVPLLTKEAVPVGLEKGEKADSGAIVAAEDGVDPV
 DGKYVYVAAKNPIIKRTHLLKFLKFLSNSGTGICNOQPLCAVDQVTTKDDVADOPARD
 REGALGVNAAVAMPYGVNEEDATIIISKLRDQATSYITIEEFLTADTILGEE
 EITRDIPVSDVLANAGEGIIIRIAEYKPDILITGKTPKSETELAHERLLRLAIF
 GEEKADVPDSEIYPPCTECPDSEVYFVRKDRLSKDELEVEAHLKDLQNGYKNO
 VATLCEVREKLGLILNNEKAPALITHRTAEIVYHEGLPEOETIERIEDEDVLDLL
 MPNCEMEVTEKGLISDTEVTLQLELTINYETVEBHIRREGDADIDHGVIRQVYVAASRR
 KLYGDKKAAGHNGKQVYSKIVPADMPYLSNGEYQOMTNPYVPSMNLGQLETH
 LGYIAKTLTGITVLYPEVGEPIVDMMIEGGLEPDKSGITLIDKTEGERDQVVG
 LTIARKLTHLADITHARSIGPISLVTOQPLSGAKMGQSGREGEWALAEYGVNIM
 LQELITVYKSDVSGRTIYVESIVKGENLRSGTPESEFVYLKEMQGLDLDVPRMYVDA
 complement(8187. .8576)
 /gene="CP0695"
 complement(8187. .8576)
 /note="similar to GB:LJ9101 SP:P41106 PID:304271
 PID:450254; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L7/L12"
 /protein_id="AAF38503.1"
 /db_xref="GI:1789609"
 /translation="MTPTSEJLTVEKLSNLVLELSQLKLEKLEMDVTAAPVAAVA
 AGGGGEAPVAAPPEPFAVTELDVPAQDKTIGLVYRVETGLAKRKAKMFTGLKTYK
 EKTSSLSDEPVTYKLDQAGKAKSAFKGL"
 complement(8620. .9132)
 /gene="CP0696"
 complement(8620. .9132)
 /note="similar to GB:I24552 SP:P41103 PID:404081;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L10"
 /protein_id="AAF38504.1"
 /db_xref="GI:1789610"
 /translation="MKRCKULLQVEEDKISAAOGFILLIRYAAASREFRSLSG
 VSAEVEYIAKTIKFKALIEAAGLEVDCSDTDGHLGVYFSGDGPVSAKQVLDFNKQHKD
 SLVPLAGMGDAASISGAIEVAVAKPLSELKRLQVVGLFAAPMSQVGVINNSYISVI
 SCVQDQAGKGN"
 complement(9148. .9846)
 /gene="CP0697"
 complement(9148. .9846)
 /gene="CP0697"
 /note="similar to SP:P04447; identified by sequence

gene	/db_xref="GI:809776" /translation="MFRVSGKKFAVVKLFKSTLTKFMSIRKTYIBCDQFSKIRID NDIVPVWRKEEYTMALALENAHVPAPRGTFENNYMSTYIGTKSGAPOLKQVEID GIYDOVIGDKRMRYAMIVHADLSEYNNMLPFRKRYFIDLAQAVDMDHMAEFLERDI VNSNFEOKHGIETDPDKIREYIKKK" complement(9745..9818) /gene="trnM" complement(9745..9818) /gene="trnM" complement(9745..9818) /note="initiator tRNA" 9843..>10142 /note="ORF100+" /codon_start=1 /transl_table=1 /protein_id="CAA48286.1" /db_xref="GI:48097" /db_xref="SWISS-PROT:003022" /translation="MAYGAVHICINOSFAGTADRSPASSPALLDPEDELPLPLK KFNQNPPLSLIDHYVCFPGVRHEITAIYGVNGVAVYGERSRQLFYFGCLPRGT"			
CDs	BASE COUNT 2926 a 2204 c 2914 g 2098 t ORIGIN			
Alignment Scores:	Pred. No.: 406 Length: 10142 Score: 91.50 Matches: 46 Percent Similarity: 37.38 Conservative: 34 Best Local Similarity: 21.50% Mismatches: 51 Query Match: 8.57% Indels: 84 Bb: 1 Gaps: 10			
US-10-024-955-7 (1-213) x TARPOG (1-10142)				
Oy 23 AsplylIeThrGlUGlUlleAsnLysAlaIleAspAspAlaIleAla-----38				
Db 6893 GAAACCTGCTCAGAGAGCTTCAGAGAGTGCCTGACATGCGCAGGCGTATGGGTTACAC 6952				
Oy 39 -----AlaIleGlUGln-----42				
Db 6953 CTGCCTGAATCGCTGCTGAGAGCATAGCTAGAGAAACAGATGGGCTGAAAAGGCC 7012				
Oy 43 -----SerGluThrIleAspProMet 49				
Db 7013 TACAGAACACTACTCAGCGGTATAGCTGAGGAGATAGGACCAATGACTGATGATCCATAC 7072				
Oy 50 LysValIProAspHisAlaAspLysPheGluThrArgHisValIleValIleValAspPheLysGly 69				
Db 7073 GAAGCC -----GTTGCAATCATAGCGGCTCAGAGC 7102				
Oy 70 -----GluLeuAlaMetArgAsnIleGluAlaValArgGlyLeuLysGln 83				
Db 7103 ATAGCGAGCCAGGTACCCAGATGACGATGAGAGACCTTCACCTTGGCGGTGTCAGGGAG 7162				
Oy 84 MetLysArg-----GlnGlyAspAlaLysValLysGlyGlnGluGlyIleValLysLysHis 102				
Db 7163 ATGAA-CGTTACTCTGGGTCTGCCAGGCTCATAGCAATGATGGA--TGCAAAGCGCAT 7218				
Oy 103 LeuLeuIleGlyValHisAspAspIle-ValSerMetLeuLys-----116				
Db 7219 ACCGAGCACTCGCTCATAGCAGATCTACTCTAGGCGCCAGATACGAGACAATGACGAGT 7278				
Oy 117 -----AspLeuAlaLysLysLeuGlyAspLeuHisProThrHisValIleSerAs 134				
Db 7279 CGTGTGATGTGTTGGAAGAGCGCTAGAGAAATACAGC-----ATAAGCGA 7323				
Oy 134 PileGlnAspPheValValAlaLeuSerLeuGluIleSerAspGlnGlyAsnIleThrMe 154				
Db 7324 TGTTCACAGATATA-----ATACTGATATAGGAGGACTTACCGT 7362				
Oy 154 Thr-----SerPheGluValArgGlnPheAl 163				
Db 7363 TACCGTGAAGCCAGATCCGAGAAAGACAAAGACAGCGCTAATCGAGATGGAAGACATAA 7422				

TITLE Extensive profiling of a complex microbial community by high-throughput sequencing
 JOURNAL Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)
 MEDLINE 22034982
 PUBMED 12039767
 REFERENCE 2 (bases 1 to 558)
 AUTHORS Hill,J.E., Seilp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G., Crosby,W.L. and Hemmingsen,S.M.
 TITLE Direct Submission
 JOURNAL Submitted (19-OCT-2001) NRC Plant Biotechnology Institute, 110 Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada
 FEATURES
 source Location/Qualifiers
 1..558
 /organism="uncultured pig faeces bacterium"
 /db_xref="taxon:190966"
 /clone="p56-0012"
 <1..>558
 /gene="cpn60"
 /gene="cpn60"
 /note="Cpn60; hsp60; groEL"
 /codon_start=1
 /transl_table=11
 /product="60 kDa chaperonin"
 /protein_id="AA049546.1"
 /db_xref="GI:21359642"
 /translation="ATVLAQAIIGVGLKNTAGANPMDLRKIDKAVAVVENTAEQAEEDGKFEKIHVAKISANSGDAIGKLAEMORKEGVITVEAKGTEFTVVEGMOFDGRTSPYEVNDETEKMECOMENPYLLIYDKISALKDLPILEPVVSGRPLLITAEIDSEALATLVNRLRGLSKICAV"
 BASE COUNT 135 a 100 c 180 g 143 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 11.4 Length: 558
 Score: 92.00 Matches: 30
 Percent Similarity: 44.92% Conservative: 23
 Best Local Similarity: 25.42% Mismatches: 31
 Query Match: 8.61% Indels: 34
 DB: 1 Gaps: 6
 US-10-024-955-7 (1-213) x AF437285 (1-558)
 Oy 13 ValAlaValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAla 32
 Db 46 GTAACCGAGGTGGCAATCCGATG-----GACTTGAAGCGTGT 84
 Oy 33 IleAspAspAlaIleAlaIle-----GluGlnSerGluThrIleAsp 47
 Db 85 ATCGACAAGGCTGTGCTGAGGTGCTGAGATATCGCTGACGACGTGAGGAGTT-- 141
 Oy 48 PrometLysValProAspHisAlaAspLysPheGluArg-----HisValGlyIleVal 65
 Db 142 -----GGCGACAAGTTCGAAAGATCGACGACTGACGAGCAATC 180
 Oy 66 AspPheLysGlyGluLeuAlaAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLys 85
 Db 181 TCTGGCATGGGACGAGGCAATCGAAAGTTGATCGCTGAGGCTATGCAGCGTGTGAAG 240
 Oy 86 ArgGlnGlyAspAlaAsnVal-----LysGlyGluGluGlyIleValLysAlaHis 102
 Db 241 AAAGAGGCTGCATCAGCTGAGAGGCTAAGGCTAGCTAGACACTACAGT----- 291
 Oy 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
 Db 292 -----GATGTCCTTGAAGGCTATGACAGTTCGACCGTGTAT 327
 RESULT 39
 TARPOT 10142 bp DNA linear BCT 03-DEC-1992
 LOCUS T.acidophilum genes rpoH, rpoB, rpoA, rpoZ, trm and ori-x,
 DEFINITION ori135, ori186 and ori10+.
 ACCESSION X68198

VERSION X68198.1 GI:48089
 KEYWORDS DNA-dependent RNA polymerase-subunit; transfer RNA-Met.
 SOURCE Thermoplasma acidophilum.
 ORGANISM Thermoplasma acidophilum
 Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 Thermoplasmataceae; Thermoplasma.
 REFERENCE
 AUTHORS Kleink,H.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-1992) H. Kleink, MPI f Biochemie, Am Klopferspitz 18, 8033 Martinsried, FRG
 REFERENCE
 AUTHORS Kleink,H.P., Renner,O., Schwass,V. and Zillig,W.
 TITLE Nucleotide sequence of the genes encoding the subunits H, B, A' and A'' of the DNA-dependent RNA polymerase and the initiator tRNA from Thermoplasma acidophilum
 JOURNAL Nucleic Acids Res. 20 (19), 5226 (1992)
 MEDLINE 93027268
 PUBMED 1408839
 FEATURES
 source Location/Qualifiers
 1..10142
 /organism="Thermoplasma acidophilum"
 /strain="DSM 1728"
 /db_xref="taxon:2303"
 26..289
 /gene="rpoH"
 26..289
 /gene="rpoH"
 /EC-number="2.7.7.6"
 /note="subunit H"
 /codon_start=1
 /transl_table=11
 /product="DNA-directed RNA polymerase"
 /protein_id="CAA48279.1"
 /db_xref="GI:48090"
 /db_xref="SWISS-PROT:Q03588"
 /translation="MAKFNVLDMVLVEHHIVPEEKEKILKEINKEFLPKISPNDPAIKALVAHGKIKKEGTIIKIVKSPMGSHVYRVVAASEVK"
 286..3873
 /gene="rpoB"
 286..3873
 /gene="rpoB"
 /EC-number="2.7.7.6"
 /note="subunit B"
 /codon_start=1
 /transl_table=11
 /product="DNA-directed RNA polymerase"
 /protein_id="CAA48280.1"
 /db_xref="GI:809775"
 /db_xref="SWISS-PROT:Q03587"
 /translation="MKELVAYFEKKYGVTHVHQLDSMNFYATPDNPNSVMOQYIDETKVSDDADGYFVLPDAKTGCHDIRIYIRGRVRENGHYVGEOTIFICKPEIKESGASNOITPNEARLIDLNTYLAIPVTLKLRIVEDGLEKSEIILKVDLPVWNSKICTISEENLDYIETKNNPGIIGLSRREKLOYGEDPDGPGYFIIIGSESEVIVSLIEDLANKINVEEDYVESKEVSEVSORGFRAITSMERKGTGNTVSPVAVGVPLVILKALGLEEDVYHDAIYAVPEMPEITYSNIEDSKNPKVLPNGVNTTDAIYSLERFAAGAKGERDOKISQMLDHSILPHLEDSPSDRKRIIYGRMARSLIELSLGIRKEDDKDHLANKIKIAGLWDELFRSAFQSVKMDLKRQLEKTYNRRKGIKIRPAVQDILLTOYVLAMSTGNNIGRTGVSQLDRVSNLSTISHLR1ISLPTROPHEARDLHPOMORICPNETPEOCNGLVKNAAALINVTGIDIDDSVMEIKGMQVRELVESPKGAVYLNCPFIYCHDPPILVYSRIEERERSGRMSDEVAVRYDDNTNEYIVSDBGRLRPLLLIKDQTYLDPTMIERLKHEISFEDVLEKOGATEMDAEEEDTVVAVYVDIPEKCPHNSYLRSMYVNPRESITTELECGCHORFNPVPSKLSKENHLEIDPMILGVASITIPYEHNSPPIITIASAMAKQSIQEFQSNVRIRPDRGLHLHPVQVLRVTRVMDIHYDRRPAQGNFVAVLSIEGVN10DALVINKAAIERGLGSTRFRTYSABERRPGQEDOKFELPTHDIIIGARAEYEVKNLDDSGIIPPEAVVEGSDVLICKTSPRELEGEERLGPQRRESSVTVRPNESGYVDNVEFLTVSESNSRVKIKVRSERIPELGDFASHHGKGVGAVYVQEDMPEEDGIIIDLIENPHSIPSRMTVGHILLEMIGKIASPTGRTIGTIFSGPEKSRARLVKYGKRSSTEWYDGTIGRRRADIFVGYIYOKLHHNVAGCFARSKGPOVILTRQPTGGRSRQVGLRFGEMERDPLIAHGAAWYIKDILLDOSGTVLYVCGNSGCHIALYDRRGTLRCPCPGCGNTGN1YIETSYAFKLMRDELISLGVVRLMDGMK"
 3870..6527
 /gene="rpoA1"

gene

Alignment Scores:			
Pred. No.:	11.4	Length:	558
Score:	92.00	Matches:	30
Percent Similarity:	44.92%	Conservative:	23
Best Local Similarity:	25.42%	Mismatch:	31
Query Match:	8.61%	Indels:	34
DB:	1	Gaps:	6

US-10-024-955-7 (1-213) x AF437078 (1-558)			
QY	13	ValAlaValSerAlaAspProIleHisTryAspLysIleHnGluGluIleAsnLysAla	32
DB	46	GTAACCGACAGGCGCAATCCGATG-----GACTGTGAAGCGCGGT	84
QY	33	IleAspAspAlaIleAlaAlaIle-----GluInSerGluThrIleAsp	47
DB	85	ATTGCAAGAGCGTGTGTAGCGTGGTGAAGATATCGGTGACAGCGTGAAGGTT--	141
QY	48	PromElytValProAspHisAlaAspLysPheGluArg-----HisValGlyIleVal	65
DB	142	-----GGCACAAGTTTCGAGCAAGATGACACATGACACCAAGATC	180
QY	66	AspPheLysGlyGluLeuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLys	85
DB	181	TCTGCGAATGGCGGACGAGCAATCGTTAAGTTGATGATCGTGAAGCGTACAGCGTGAAG	240
QY	86	ArgGlnGluAspAlaAsnVal-----LysGlyGluGluGlyIleValLysAlaHis	102
DB	241	AAAGAGGCGTCACTACCTGTGTGAGGAGCGTAAAGGCTGACTGACATCAATCAATG-----	291
QY	103	LeuLeuIleGlyValHisAspAspIleValSerMetGluTryAspLeuAlaTyr	120
DB	292	-----GATGTCGTTGAAGTATGACAGTTGACCGTGGTAT	327

RESULT 35			
LOCUS	AF437182	558 bp	DNA linear BCT 09-JUN-2002
DEFINITION	Uncultured pig faeces bacterium clone 016_B02 60 kDa chaperonin (cpn60) gene, partial cds.		
ACCESSION	AF437182		
VERSION	AF437182.1	GI:21359435	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			

```

1. .558
/organism="uncultured pig faeces bacterium"
/db_xref="taxon:190966"
/clone="016_B02"
<1..>558
/gene="cpn60"
<1..>558
/gene="cpn60"
/note="cpn60; hsp60; groEL"
/codon_start=1
/transl_table=11
/product="60 kDa chaperonin"
```

BASE COUNT		134 a	103 c	179 g	142 t
ORIGIN					
Alignment Scores:					
Pred. No.:	11.4	Length:	558		
Score:	92.00	Matches:	30		
Percent Similarity:	44.92%	Conservative:	23		
Best Local Similarity:	25.42%	Mismatches:	31		
Query Match:	8.61%	Indels:	34		
DB:	1	Gaps:	6		
US-10-024-955-7 (1-213) x AF437182 (1-558)					
QY	13	VALAIAValserAlaaspProIleHisIyrAspIysIleIthngIugIuIleAsnIysAla	32		
DB	46	GRAACCGCAGGTGCGAACCCGATG-----GACTTGAAGCGCGT	84		
QY	33	IleAspAspAlaIleAlaIle-----GluGlnSerGIuThrIleAsp	47		
DB	85	ATTGACAAAGCGCTGTGGCTTAGGTGCTCGAATATCGCTGACAGCGTGAAGAGGTT---	141		
QY	48	PrometIysValProAspHisAlaAspIysPheGluArg-----HisValGIlyIleVal	65		
DB	142	-----CGCGACAAGTTCGAGAAGATGACACATGTAGCGCAAGATC	180		
QY	66	AspPheIysGIyGIuLeuAlaMetkArgAsnIleGluAlaArgGIyLeuIysGlnMetIys	85		
DB	181	TCTGGCAATGTGGCGACAGCAATCGAAATGTGATCGCTGAGCGTACAGCTGTGAAG	240		
QY	86	ArgGInGIyAspAlaAsnVal-----LysGIyGluGluGIyIleValIysAlaHis	102		
DB	241	AAAGAGGGGTGTCATCAGCTTTGAGAGCGTAAAGGTACTAGACACTACAGTG-----	291		
QY	103	IeuLeuIleGIyAlaHisAspIleValSerMetGIuIyrIysLeuAlaIatyr	120		
DB	292	-----GATGTCTGTGAGGGTATGACAGTTCGACCGTGTAT	327		
RESULT 36					
LOCUS	AF437184	558 bp	DNA	linear	BCT 09-JUN-2002
DEFINITION	Uncultured pig faeces bacterium clone 016_C09 60 kda chaperonin (cpn60) gene, partial cds.				
ACCESSION	AF437184				
VERSION	AF437184.1	GI:21359439			
KEYWORDS	.				
SOURCE	uncultured pig faeces bacterium.				
ORGANISM	uncultured pig faeces bacterium				
REFERENCE	Bacteria: environmental samples.				
AUTHORS	1 (bases 1 to 558)				
TITLE	Hill,J.E., Seilp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G., Crosby,W.L. and Hemmingsen,S.M.				
JOURNAL	Extensive profiling of a complex microbial community by high-throughput sequencing				
MEDLINE	Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)				
PUBMED	22034982				
REFERENCE	12039767				
AUTHORS	2 (bases 1 to 558)				
TITLE	Hill,J.E., Seilp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G., Crosby,W.L. and Hemmingsen,S.M.				
JOURNAL	Direct Submission				
FEATURES	Submitted (19-Oct-2001) NRC Plant Biotechnology Institute, 110 Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada				
SOURCE	Location/Qualifiers				
	1..558				
	/organism="uncultured pig faeces bacterium"				
	/db_xref="taxon:190966"				
	/clone="016_C09"				


```

Db      85  ATTGACAAGCCTGCTGTAAGGTGTCGACAAATATCCCTGACGAGCGCTGAGAGGTC--- 141
Oy      48  PrometlyValProaspHisAlaAspLysPheGluArg-----HisValGlyLeval 65
Db      142  -----GGCGACAACTGTGAGAGATTCGACGATGAGCAAGATC 180
Oy      66  AspPheLysGlyLuleuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLys 85
Db      181  TCTGCGAATGCGACGAGCAATCGAAAGTTGATGCTGAGCGTATGCAAGCGTGTAAG 240
Oy      86  ArgGlnGlyAspAlaAsnVal-----LysGlyGluGluGlyLevalLysAlaHis 102
Db      241  AAGAGGCTGTCACTGCTTGAAGAGCTAAGGCTACTGAGCTACAGCTG----- 291
Oy      103  LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
Db      292  -----GATGTCGTTGAGGCGATGCGAGTTCGACCGCTGTTAT 327

RESULT 33
AF436980      558 bp      DNA      linear      BCT 09-JUN-2002
LOCUS      uncultured pig faeces bacterium clone 004_B01 60 kda chaparonin
DEFINITION      (cpn60) gene, partial cds.
ACCESSION      AF436980
VERSION      AF436980.1 GI:21359031
KEYWORDS
SOURCE      uncultured pig faeces bacterium.
ORGANISM      Bacteria; environmental samples.
REFERENCE      1 (bases 1 to 558)
AUTHORS      Hill,J.E., Seilp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G.,
              Crosby,W.L. and Hemmingsen,S.M.
              Extensive profiling of a complex microbial community by
              high-throughput sequencing
              Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)
PUBMED      12039767
JOURNAL      2 (bases 1 to 558)
MEDLINE      Hill,J.E., Seilp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G.,
              Crosby,W.L. and Hemmingsen,S.M.
              Direct Submision
              Submitted (19-OCT-2001) NRC Plant Biotechnology Institute, 110
              Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada
FEATURES
     source
         1..558
         /organism="uncultured pig faeces bacterium"
         /db_xref="taxon:190966"
         /clone="004_B01"
         <1..>558
         /gene="cpn60"
         <1..>558
         /gene="cpn60"
         /note="cpn60; hsp60; groEL"
         /codon_start=1
         /transl_table=11
         /product="60 kda chaparonin"
         /protein_id="AA049241.1"
         /db_xref="GI:21359032"
         /translation="ATVLAQAIIGVGLKNTVAGANDMDLKRIGDKAVARVENIAEQA
         EEVGDKEFEIENHAKISANGDEAIKGLIEAMQRYVKEGVITVEAKGSETVDVEG
         MOPDRGIVISPYFTDEKMECOMENYIILYDKKISALMDLLPILPVPVQSGRPLLI
         AEDIDSEALATLVNLRSLKRCAY"
BASE COUNT      135 a      100 c      178 g      145 t
ORIGIN
Alignment Scores:
Pred. No.:      11.4      Length:      558
Score:      92.00      Matches:      30
Percent Similarity:      44.92%      Conservative:      23
Best Local Similarity:      25.42%      Mismatches:      31
Query Match:      8.61%      Indels:      34
DB:      1      Gaps:      6

```

```

US-10-024-955-7 (1-213) x AF436980 (1-558)
Oy      13  ValAlaValSerLlaAspProIleHisTyrAspLysIleThrGluIleAsnLysAla 32
Db      46  GTTACCCGACGCTGCGAATCCGATG-----GACTTGAAGCGGT 84
Oy      33  IleAspAspAlaIleAlaIle-----GluIleSerGluThrIleAsp 47
Db      85  ATGACAAGCCTGCTGTAAGGTGTCGACAAATATCCCTGACGAGCGTGAGAGTT--- 141
Oy      48  PrometlyValProaspHisAlaAspLysPheGluArg-----HisValGlyLeval 65
Db      142  -----GGCGACAACTGTGAGAGATTCGACGATGAGCAAGATC 180
Oy      66  AspPheLysGlyLuleuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLys 85
Db      181  TCTGCGAATGCGACGAGCAATCGAATGCTGATGCTGAGCGTATGCAAGCGTGTAAG 240
Oy      86  ArgGlnGlyAspAlaAsnVal-----LysGlyGluGluGlyLevalLysAlaHis 102
Db      241  AAGAGGCTGTCACTGCTTGAAGAGCTAAGGCTACTGAGCTACAGCTG----- 291
Oy      103  LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
Db      292  -----GATGTCGTTGAGGCGATGCGAGTTCGACCGCTGTTAT 327

RESULT 34
AF437078      558 bp      DNA      linear      BCT 09-JUN-2002
LOCUS      uncultured pig faeces bacterium clone 008_H04 60 kda chaparonin
DEFINITION      (cpn60) gene, partial cds.
ACCESSION      AF437078
VERSION      AF437078.1 GI:21359227
KEYWORDS
SOURCE      uncultured pig faeces bacterium.
ORGANISM      Bacteria; environmental samples.
REFERENCE      1 (bases 1 to 558)
AUTHORS      Hill,J.E., Seilp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G.,
              Crosby,W.L. and Hemmingsen,S.M.
              Extensive profiling of a complex microbial community by
              high-throughput sequencing
              Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)
PUBMED      12039767
JOURNAL      2 (bases 1 to 558)
MEDLINE      Hill,J.E., Seilp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G.,
              Crosby,W.L. and Hemmingsen,S.M.
              Direct Submision
              Submitted (19-OCT-2001) NRC Plant Biotechnology Institute, 110
              Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada
FEATURES
     source
         1..558
         /organism="uncultured pig faeces bacterium"
         /db_xref="taxon:190966"
         /clone="008_H04"
         <1..>558
         /gene="cpn60"
         <1..>558
         /gene="cpn60"
         /note="cpn60; hsp60; groEL"
         /codon_start=1
         /transl_table=11
         /product="60 kda chaparonin"
         /protein_id="AA049339.1"
         /db_xref="GI:21359228"
         /translation="ATVLAQAIIGVGLKNTVAGANDMDLKRIGDKAVARVENIAEQA
         EEVGDKEFEIENHAKISANGDEAIKGLIEAMQRYVKEGVITVEAKGSETVDVEG
         MOPDRGIVISPYFTDEKMECOMENYIILYDKKISALMDLLPILPVPVQSGRPLLI
         AEDIDSEALATLVNLRSLKRCAY"
BASE COUNT      135 a      101 c      178 g      144 t
ORIGIN

```

Db	292	-----		:::			
RESULT 31	AF436950	558 bp	DNA	linear	BCT 09-JUN-2002		
LOCUS	AF436950						
DEFINITION	Uncultured pig faeces bacterium clone 003_02A 60 kda chaperonin (cpn60) gene, partial cds.						
ACCESSION	AF436950						
VERSION	AF436950.1	GI:21358971					
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
MEDLINE							
PUBMED							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							
gene							
CDS							
BASE COUNT							
ORIGIN							
Alignment Scores:							
Pred. No.:							
Score:							
Percent Similarity:							
Best Local Similarity:							
Query Match:							
DB:							
US-10-024-955-7 (1-213) x AF436950 (1-558)							
QY	13	VALAIAValSerAlaaspProIleHisTryAspLysIleThrGluGluIleasnLysAla	32				
Db	46	GTAACCGCAGAGTGCACAAATCCGATG-----GACTTGAAGCGTGCT	84				
QY	33	IleAspAspAlaIleAlaAlaIle-----GluGlnSerGluThrIleAsp	47				
Db	85	ATCGACAAAGCGCTGTGGCTTAAGTGTGCGAATATCCGCGACGAGCGTATGAGAGCTT---	141				
QY	48	PrometLysValProAspHisAlaAspLysPheGluArg-----HisValGlyIleVal	65				
Db	142	-----GGCGACCAAGTTCGAGGAAGATCGACAGATGTATACCAAGAATC	180				
QY	66	AspPheLysGlyGluIleuAlaMetcArgAsnIleGluAlaArgGlyLeuLysGlnMetLys	85				

Db	181	TCTGGATGATGCCAGACGAGCATATCGGTACTGCATATCACCCTGTGAAG	240
Oy	86	ArgInGIYAspAlaAsnVal-----LysGLYgluIguIGLYllyleValAlaHis	102
Db	241	AAGAAGGTGTCAATCACCTGTGAGAGAGCTAAAGCGTACTGACTAGTACAGTG-----	291
Oy	103	LeuLeuIIeGIYAlaHIsAspAlleValserMetGLUTyrAspLeuAlaTyr	120
Db	292	-----GATGTGCTTTAAGSGTAGTCAGTTCGACCGCTGTATT	327
RESULT 32	AF436963	558 bp	DNA linear BCT 09-JUN-2002
LOCUS	AF436963		
DEFINITION	Uncultured pig faeces bacterium clone 003_D03_60 kDa chaperonin (cpn60) gene, partial cds.		
ACCESSION	AF436963		
VERSION	AF436963.1	GI:21358997	
KEYWORDS	.		
SOURCE	uncultured pig faeces bacterium.		
ORGANISM	Bacteria; environmental samples.		
REFERENCE	1 (bases 1 to 558)		
AUTHORS	Hill,J.E., Seip,R.P., Betts,M., Hawkins,L., Van Kessel,A.G., Crosby,W.L. and Hemmingsen,S.M.		
TITLE	Extensive profiling of a complex microbial community by high-throughput sequencing		
JOURNAL	Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)		
MEDLINE	22034982		
PUBMED	12039767		
REFERENCE	2 (bases 1 to 558)		
AUTHORS	Hill,J.E., Seip,R.P., Betts,M., Hawkins,L., Van Kessel,A.G., Crosby,W.L. and Hemmingsen,S.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-OCT-2001) NRC Plant Biotechnology Institute, 110 Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada		
FEATURES	Location/Qualifiers		
source	1..558		
gene	/organism="uncultured pig faeces bacterium"		
CDS	/db_xref="taxon:190966" /clone="003_D03" <1..>558 /gene="cpn60" /length="558" /note="Cpn60; hsp60; groEL" /codon_start=1 /transl_table=1 /product="60 kDa chaperonin" /protein_id="AAA49224.1" /db_xref="GI:21358997" /translation="AVVLAQAIIIVGAKNNVTAGANPMDLRKIDKAAYKVVENIAEQAEVDQDFKEIEHVAKISANGDEAIKGLIAEMQMVKREGVTVTEAGTETTVVVGSA MDPGRGISPEFVDTMECOMENPYLLDYDKRISALKDPLPLEPVVSGRRLLIIIT AEDIDSEALATLVNRLRGSLICAV"		
BASE COUNT	135 a 100 c 179 g 144 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	11.4	Length:	558
Score:	92.00	Matches:	30
Percent Similarity:	44.92%	Conservative:	23
Best Local Similarity:	25.42%	Mismatch:	31
Query Match:	8.61%	Indels:	34
DB:	1	Gaps:	6
US-10-024-955-7 (1-213) x AF436963 (1-558)			
Oy	13	valalavalseralaspProIleHisTYrAspLyIleThrGlulIeasnlYsala	32
Db	46	GTAACCGAGAGTCCGCAATCCGATG-----GACTTGAAAGCGTGT	84
Oy	33	IleaspsapAlaIleAlaAlaIle-----GluGlnSerGIUThrIleasp	47

TITLE Crosby, W.L. and Hemmingsen, S.M.
Extensive profiling of a complex microbial community by
high-throughput sequencing
JOURNAL Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)
MEDLINE 22034982
PUBMED 12039767
REFERENCE 2 (bases 1 to 558)
AUTHORS Hill, J.E., Seilp, R.P., Betts, M., Hawkins, L., Van Kessel, A.G.,
Crosby, W.L. and Hemmingsen, S.M.
TITLE Direct Submission
JOURNAL Submitted (19-Oct-2001) NRC Plant Biotechnology Institute, 110
Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada
FEATURES
source Location/Qualifiers
1..558
/organism="uncultured pig faeces bacterium"
/db_xref="taxon:190966"
/clone="002_C04"
<1..>558
/gene="cpn60"
/note="cpn60"
/gene="cpn60"
/note="cpn60; hsp60; groEL"
/codon_start=1
/transl_table=11
/product="60 kDa chaperonin"
/protein_id="AAM49190.1"
/db_xref="GI:21358930"
/translation="ATVLAQALITGVKNTAGAMPMDKRGIDKAVAKVENIAEQA
EEVGKFEKIEHVAKISANGDEAIKGLIAEAQRYKKEVITVEAKGETVDVVEG
KQFDGKISPYFVDTKEKMECMENPYIILYDKKISALKDLPILEPVVQSRPLLI
AEDISEALATLVNRLRSLKICAV"
BASE COUNT 135 a 103 c 178 g 142 t
ORIGIN
Alignment Scores:
Pred. No.: 11.4 Length: 558
Score: 92.00 Matches: 30
Percent Similarity: 44.92% Conservative: 23
Best Local Similarity: 25.42% Mismatches: 31
Query Match: 8.61% Indels: 34
DB: 1 Gaps: 6
US-10-024-955-7 (1-213) x AF436929 (1-558)
QY 13 ValAlaValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAla 32
Db 46 GTAACCGCAGCTCGAATCCGATG-----GACTTGAAGCGCGGT 84
QY 33 IleAspAspAlaIleAlaIle-----GluGlnSerGluThrIleAsp 47
Db 85 ATTGACAAGCGCTGCTAAGTGTGCGAATATCGCTGAGCAGCTGAGGAGTT--- 141
QY 48 PrometLysValProAspHisAlaAspLysPheGluArg-----HisValGlyIleVal 65
Db 142 -----GGCGACAAGTTGCGACAAGATCGACATGATGAGGAGATC 180
QY 66 AspPheLysGlyLysLeuAlaMetArgAsnIleGluAlaArgLysGlyLeuLysGlnMetLys 85
Db 181 TCTGCAATGCGCAGCAATCGAATGCTGATGCTGAGCTATGCGCTGTGAAG 240
QY 86 ArgGlnGlyAspAlaAsnVal-----LysGlyGluGluGlyIleValLysAlaHis 102
Db 241 AAAGAGGCTGTCTACTCTGTGAGAGCGTAAGGTAAGTACTGACACTACAGG----- 291
QY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
Db 292 -----GATGCTGTGAGAGGATGATGACAGTTCACCGCTGTAT 327
RESULT 30
AF436936 558 bp DNA linear BCT 09-JUN-2002
LOCUS uncultured pig faeces bacterium clone 002_E07 60 kDa chaperonin
DEFINITION (cpn60) gene, partial cds.

ACCESSION AF436936
VERSION AF436936.1 GI:21358943
KEYWORDS
SOURCE uncultured pig faeces bacterium.
ORGANISM Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 558)
AUTHORS Hill, J.E., Seilp, R.P., Betts, M., Hawkins, L., Van Kessel, A.G.,
Crosby, W.L. and Hemmingsen, S.M.
TITLE Extensive profiling of a complex microbial community by
high-throughput sequencing
JOURNAL Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)
MEDLINE 22034982
PUBMED 12039767
REFERENCE 2 (bases 1 to 558)
AUTHORS Hill, J.E., Seilp, R.P., Betts, M., Hawkins, L., Van Kessel, A.G.,
Crosby, W.L. and Hemmingsen, S.M.
TITLE Direct Submission
JOURNAL Submitted (19-Oct-2001) NRC Plant Biotechnology Institute, 110
Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada
FEATURES
source Location/Qualifiers
1..558
/organism="uncultured pig faeces bacterium"
/db_xref="taxon:190966"
/clone="002_E07"
<1..>558
/gene="cpn60"
/note="cpn60"
/gene="cpn60"
/note="cpn60; hsp60; groEL"
/codon_start=1
/transl_table=11
/product="60 kDa chaperonin"
/protein_id="AAM49197.1"
/db_xref="GI:21358944"
/translation="ATVLAQALITGVKNTAGAMPMDKRGIDKAVAKVENIAEQA
EEVGKFEKIEHVAKISANGDEAIKGLIAEAQRYKKEVITVEAKGETVDVVEG
KQFDGKISPYFVDTKEKMECMENPYIILYDKKISALKDLPILEPVVQSRPLLI
AEDISEALATLVNRLRSLKICAV"
BASE COUNT 135 a 102 c 179 g 142 t
ORIGIN
Alignment Scores:
Pred. No.: 11.4 Length: 558
Score: 92.00 Matches: 30
Percent Similarity: 44.92% Conservative: 23
Best Local Similarity: 25.42% Mismatches: 31
Query Match: 8.61% Indels: 34
DB: 1 Gaps: 6
US-10-024-955-7 (1-213) x AF436936 (1-558)
QY 13 ValAlaValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAla 32
Db 46 GTAACCGCAGCTCGAATCCGATG-----GACTTGAAGCGCGGT 84
QY 33 IleAspAspAlaIleAlaIle-----GluGlnSerGluThrIleAsp 47
Db 85 ATTGACAAGCGCTGCTAAGTGTGCGAATATCGCTGAGCAGCTGAGGAGTT--- 141
QY 48 PrometLysValProAspHisAlaAspLysPheGluArg-----HisValGlyIleVal 65
Db 142 -----GGCGACAAGTTGCGACAAGATCGACATGATGAGGAGATC 180
QY 66 AspPheLysGlyLysLeuAlaMetArgAsnIleGluAlaArgLysGlyLeuLysGlnMetLys 85
Db 181 TCTGCAATGCGCAGCAATGCTGATGCTGAGCTATGCGCTGTGAAG 240
QY 86 ArgGlnGlyAspAlaAsnVal-----LysGlyGluGluGlyIleValLysAlaHis 102
Db 241 AAAGAGGCTGTCTACTCTGTGAGAGCGTAAGGTAAGTACTGACACTACAGG----- 291
QY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120

```

QY 14 AlaValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIle 33
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
Db 66113 ACTCTTAGAGCTTCTCCCTCCCTCGTTGACTTAACGACCGAAGCTCAACGAGTAATT 66172
QY 34 AspAspAlaIleAlaIleAlaIleGlu-----GlnSerGluThrIleAspPromet 49
      ||| ::::
Db 66173 GATTTAGTGGTGGCGGACGCTGTAAGTACGAGTCCAACTGTAACAAATGATTAACGG 66232
QY 50 LysVal-----Pro 52
      |||
Db 66233 GCCGTGAGCTCCATTAAATAGATCCTTGCTGTTTGTGACAGAGTCCGCCGATCCG 66292
QY 53 AspHisAlaAspLysPheGluArgHisValGlyIle----- 64
      ||| ::::::
Db 66293 GATGAAATTAACAGGATCTTGATGATCAGTACGAGGATCAAGAAATGGGAGATTGCCGTTAA 66352
QY 65 ---ValAspPheLysGlyLysIleuAlaMetArgAsnIleGluAlaArgGlyLeuLysGln 83
      ::::::::::: ||| ::::
Db 66353 GAGATTAACCTTCTCGGAAA-----GAGATGATGTCAGAACGCCGTTAGATGAG 66400
QY 84 MetLysArgGlnGlyAspAlaAsnValLysGlyGluGlyIleValLysAlaHisLeu 103
      ||| ::::
Db 66401 AAGAAAG-----AAGAGTACGATCAAGTCCAGCTTCTCAGTAGT 66442
QY 104 LeuIleGlyVal----- 107
      ||| ::::
Db 66443 TTAATGCGATTCATATTGTCAGATGATTAATGAGACGATGACGAGATGAT 66502
QY 108 -----HisAspAspIleValSerMetGluTyrAspLeuAlaTyrLysLeuGly 123
      ||| ::::
Db 66503 CATCATTAATCATCATGAAATGATGATTAAGAAAGATCATAGTATGCCAGGATTTG--- 66559
QY 124 AspLeuHisProThrThrHisValIleSerAspIleGluAspPheValAlaIleAsp 143
      ::::::::::: :::::::::::
Db 66560 -----AAGTTGAGAGATGCTCTTTGTCCTCAATTTCA 66589
QY 144 LeuGluIleSerAspGluLysAsnIleThrMetThrSerPheGluValArgGlnPheAla 163
      ||| ::::
Db 66590 CTAGAGATTAATGACGATCCTGTTGTTATGAACAGCGCACCTTATGATCGG---AGC 66646
QY 164 AsnValValAsnHisIleGlyLysSerIleLeuAspProIlePheGlyValLeuSer 183
      ::::::::::: ||| ::::
Db 66647 TCCATCTACTAAATGGTTCGATCCGCGAACATCACCTGCCCTATTAACCGGAAGATTCTG 66706
QY 184 AspValLeuThrAlaIlePheGlnAspThrValArgLysGluMetThrLys----- 200
      ::::
Db 66707 ACTAGTACCGAGTTGGTTGATTAACGTATCCGTGAGGCAAGTGATTCGAAAGACCTCAAA 66766
QY 201 -----ValLeuAlaProAlaPheLysArg 208
      ||| ::::
Db 66767 ACAACGGCATGCTTTTGGCGGGGATTATGCGGAGA 66802

RESULT 28
AF436928 558 bp DNA linear BCT 09-JUN-2002
LOCUS Uncultured pig faeces bacterium clone 002_C03 60 kDa chaperonin
DEFINITION (cpn60) gene, partial cds.
ACCESSION AF436928
VERSION AF436928.1 GI:21358927
KEYWORDS
SOURCE
ORGANISM uncultured pig faeces bacterium.
Bacteria: environmental samples.
REFERENCE
AUTHORS Hill,J.E., Seipp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G.,
TITLE Extensive profiling of a complex microbial community by
high-throughput sequencing
JOURNAL Appl. Environ. Microbiol. 68 (6), 3055-3066 (2002)
MEDLINE 22034982
PUBMED 12039767
REFERENCE 2 (bases 1 to 558)
AUTHORS Hill,J.E., Seipp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G.,

```

```

Crosby,W.L. and Hemmingsen,S.M.
Direct Submission
Submitted (19-Oct-2001) NRC Plant Biotechnology Institute, 110
Gymnasium Place, Saskatoon, Saskatchewan S7N 0W9, Canada
FEATURES
source
1..558
/organism="uncultured pig faeces bacterium"
/db_xref="taxon:190966"
/clone="002_C03"
<1..>558
/gene="cpn60"
<1..>558
/gene="cpn60"
/feature="cpn60"
/feature="hsr60; groEL"
/codon_start=1
/transl_table=11
/product="60 kDa chaperonin"
/protein_id="AA049189.1"
/db_xref="GI:21358928"
/translation="ATVLAQALIGVGLKNVTAGANPMDLRGIDRAVAKVENIAEOA
FEVGRKFEIHHVAKISANGDEAIKLIEMARVYKKEGVTEAEAKGTEYDVVEG
MQRDGYISPVVPTDEKMECOMENDYILITDKKISALKDLPLIEPVVQSGRPILLII
AEDIDSEALATLVNRLRSLKICAY"
BASE COUNT 134 a 100 c 179 g 145 t
ORIGIN

Alignment Scores:
pred. No.: 11.4 length: 558
Score: 92.00 Matches: 30
Percent Similarity: 44.92% Conservative: 23
Best local Similarity: 25.42% Mismatches: 31
Query Match: 8.61% Indels: 34
DB: 1 Gaps: 6

US-10-024-955-7 (1-213) x AF436928 (1-558)
QY 13 ValAlaValSerAlaAspProIleHisTyrAspLysIleThrGluGluIleAsnLysAlaIle 32
      ||| :::::::::::
Db 46 GTAGCGGAGGTGCGAATCGGATG-----GACTTGAAAGCCTGGT 84
QY 33 IleAspAspAlaIleAlaIleAlaIle-----GluGlnSerGluThrIleAsp 47
      ||| ::::
Db 85 ATCGACAGAGCTGTGCTTAAGGTGTCGAGAAATATGCTGACGACGCTGAGGAGTT--- 141
QY 48 ProMetLysValProAspHisAlaAspLysPheGluArg-----HisValGlyIleVal 65
      :::::::::::
Db 142 -----GGCGACAGATTGGAAGAAGATCGACGATCGAGCAAGATC 180
QY 66 AspPheLysGlyLysIleuAlaMetArgAsnIleGluAlaArgGlyLeuLysGlnMetLys 85
      ||| ::::
Db 181 TCTGCCAATGGCGACGACGCAATCGTAACTGATGCTGAGGCTATGACGCGTGGAG 240
QY 86 ArgGlnGlyAspAlaAsnVal-----LysGlyGluGluGlyIleValLysAlaHis 102
      ::::::::::: ||| ::::
Db 241 AAGAGAGGTGTCATCACTGTTGAGAGGCTAAAGGCTACGACACTACAGT----- 291
QY 103 LeuLeuIleGlyValHisAspAspIleValSerMetGluTyrAspLeuAlaTyr 120
      ||| ::::
Db 292 -----GATGTCTTGAGGATGATGATGATGATGATGATGATGATGAT 327

RESULT 29
AF436929 558 bp DNA linear BCT 09-JUN-2002
LOCUS Uncultured pig faeces bacterium clone 002_C04 60 kDa chaperonin
DEFINITION (cpn60) gene, partial cds.
ACCESSION AF436929
VERSION AF436929.1 GI:21358929
KEYWORDS
SOURCE
ORGANISM uncultured pig faeces bacterium.
Bacteria: environmental samples.
REFERENCE
AUTHORS Hill,J.E., Seipp,R.P., Betts,M., Hawkins,L., Van Kessel,A.G.,

```

CDS
 /protein_id="AAD39561.1"
 /db_xref="GI:5091532"
 /translation="MGCGVTPKSCVGAKLRSSKRRKSRNRRAKIQKRRAVSSRLSDG
 SPFDLDHHRNSNPSRGLSLRLSVSSAPD"
 complement(join(5929..6152,6393..6466,6659..6726,
 6855..6932,7008..7297,7466..7502))
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="t10024.2"
 /protein_id="AAD39562.1"
 /db_xref="GI:5091533"
 /translation="MDHVSSTKVEFENIRALEERLKCHETELQSKDNIIISLAOLE
 SEKSKNEHQHVEEISLSELACTAFSKKLVDHFSRICLDDQYLOKTYQDAVE
 NLISKEALYSEMGKLEMLIORIOESLMPEDRKRFTSLTFPGOSPERKRRBN
 DTVDKMEELVYAPVMSHOENSVKVPSPSPRCOKTDCRSIQDDHQLDSAEVLQH
 NTVSGHSTNNHNFLELASIDFLFVSCLOLLIEVK"
 complement(join(8063..8221,8301..8543))
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="t10024.3"
 /protein_id="AAD39563.1"
 /db_xref="GI:5091534"
 /translation="MEMLKSKFKLQLOSMIGEVDLRVRRRRLFYIGLCIFAPPV
 SLFEFFVFHFPTGVFAGKAIHSRSHVSCSGSSINKKQTEEDCIRKVELQADLA
 SSREFOALERKVKLLFECSISGVSESP"
 complement(join(9791..9964,10039..10230,10415..10492,
 10572..10660,10800..10914,11271..11523))
 /note="similar to protein phosphatase type 2A
 (emb|CA807806); similar to ESTs emb|235739, gb|AA869861,
 gb|AA490696, gb|AA791751, and gb|AA042146"
 /codon_start=1
 /evidence=not_experimental
 /product="t10024.4"
 /protein_id="AAD39564.1"
 /db_xref="GI:5091535"
 /translation="MPSNGELDRQIEQLMECKPLSEADVPTLCDQARALVEEYVNPQ
 VKPVTVCADHIGOFYDLIELDRIGGNADPTNYLPMGDVDRGYSEVETSLVALKV
 RYRDLRTLRGNHESROITQVGYFDECLRKGANVMKFTFDLFDYLPRLTISQV
 PCLHGCLSPSLDPTLDIRISLDRIOEPHEGPKCDLMSPPDRCCMGISPRACITGTC
 ODIAQFNNHNLISLISRAHQLMVEGFNMCQDKNVVTVFSAPNCTRCGNMAALIEIG
 ENMEQNFLOEDAPRQVEPDTTRKTPDYFL"
 complement(12271..12342)
 /product="tRNA-Pro"
 /note="codon recognized: CCU"
 /evidence=not_experimental
 /product="t10024.5"
 /protein_id="AAD39565.1"
 /db_xref="GI:5091536"
 /translation="MRDLITGTPISYEMICLFPRLVMPGFIKASFCSSRNSAG
 ESGRLKKRAREDVHTDOKRKPEVSSKRGELNKLPRITDALTYLAKAVDITHNKEK
 YESFLMEKFAQITDNGVIERIKVLEKGRDILNLTGNTFLPGKYITLLPEEKH
 KILRVFKDAIFGVTKIKTRFGDDEHAKRFEDLILNLYRKKKSISEVEEVLMLPEKH
 EDLMEFVNFILPNCESAPSTKANVRHKGATTTAMHSDKRRKORCLBDEYSGHSDR
 EDGENLVTCGAGNSLEFGLPVLRSMLIFADSPVGEQPGYPRDYERREDTEGTA
 DRTRESAASGODIGNHSGTQVGTPINELDSCOTCOTSPYRLPKRYAVEITSYR
 NTLGKKTLDLHVSVTSGSEDSFSHMKRNOYEBESLPREDRREYEDMLIGYSNAIK
 OVEILERKNNNTISVSTICIEKLHSANNLKIEKLVDNCLDPMVLLKMMHSLAP
 VILTRIKQOEEMARCHSDPQVMAVYAKNNHKSIDHSEFQKDDSNLSTKLCLVA
 EVKDISSEKHQEDLOALAVRYMPLTFPLEDFNYCDOQHEDLYLKYCECEIATE
 OSDKYMKLWITFLEPLFGLISRODNLAEVSEKNNLEODACLAVKETSAGSKRK
 HPISPKRLSKNTKMGSSSRSDVSAKIKVTKAODPKLODDAAMTNEVIOSSKFPVSPK
 NQJIMDEGNNHVNASVEKHELBEGEISLPTASRQDSNEFVGQNAFKRLQQLVTDNVR
 SNKDKSCDKGAKNKTAEADKQENCHKLSENNKTASEMLVSGTKVSCHEENRVMN

CDS
 /protein_id="AAD39566.1"
 /db_xref="GI:5091537"
 /translation="MAQKNLNLIELSQYEDDPVVLKRLSDSLYYSAQLYLPKQ
 EMEHLEPMHDILVRKLAGAEVVRVDSVDVDTYRLAKRNOQYLLGQWGLIKN
 AKVLTGSHGIGFMKDLREYERKHFKQSGITMHRACGTTSQKVVLDKK"
 complement(20664..21317)
 /note="similar to germ1-like protein (gi|2935521);
 similar to ESTs gb|T88481 and gb|A1099566"
 /codon_start=1
 /evidence=not_experimental
 /product="t10024.7"
 /protein_id="AAD39567.1"
 /db_xref="GI:5091538"
 /translation="MLNLTLLMGRVNSDDPLDQVSPSPSSHQOIFLNGKL
 CKDPQASVSDFSTALSRLPQNTKTRPPMINVYTTANLPCLNTVGLTMRDLGGS
 GVPPHVPRASEVFTCLDGYLVGLGVFTSGVFPQELHPCGTFEYFPGGLHIFLYNID
 TVSSALAVGSLSSONGTOIVSLSFISKPPPLVLSAVDINODVARIKSLSEG"
 complement(join(23710..24085,24185..24255,24338..24415,
 24503..24579,24668..24654))
 /note="similar to response regulator 1 (dbj|BA311431);
 similar to ESTs gb|T43772, emb|234204, gb|AA067391, and
 emb|234620"
 /codon_start=1
 /evidence=not_experimental
 /product="t10024.8"
 /protein_id="AAD39568.1"
 /db_xref="GI:5091539"
 /translation="MARDGVCLRRSEKMSYGGIGIESAPLDLDEHVLAVDDSLV
 DRIVERLLRTISCVTAVDSCMALFEFLGDNESAFAFDKLYDLITDVCMTG
 GYELIKIKESNPREVPVYIMSSFNVLTRIDRCLEKADQLKPVRLADVKKRASH
 LTKDKVLSGNKRKRLPEDSSVNSLSLPPSPPLTTPSPSSPPLTVTSTSSDSPLSP
 VEFTSPSLSSIDDEDVLTSSSESPTRKRRSPELD"
 complement(31455..31455)
 /note="putative zinc finger protein 5 (pir|S55885)"
 /codon_start=1
 /evidence=not_experimental
 /product="t10024.9"
 /protein_id="AAD39569.1"
 /db_xref="GI:5091540"
 /translation="MSINPTMSRTGESSGSSSDTKILFGFELISGRTPETITAES
 VSSNTNTSLTYMKRHECOYCEKAFNSALGHNARKKELKRLLOLORRASIG
 YLTNHOPIRTSPOROYKTPSYCAFPSSMHNVDQMGYNEDMSSRSOINGNNDTC
 ODLPQSGEMKLYVRPMIDOPQDLSRSQMSINSLDHLFPADAA"
 join(34675..34779,34868..34959,35119..35370,35452..35574,
 35654..35827,35940..36050,36130..36263,36349..36442,
 36641..36739,36824..36975,37035..37180,37669..37792,
 37887..38001,38055..38154,38234..38268,38510..38686,
 38770..38807,38875..38968,39046..39225,39319..39433,
 39530..39599,39729..39876,40074..40202,40333..40458,
 40612..40677,40772..41011)
 /note="unknown; similar to ESTs dbj|C26360, gb|T45760, and

Alignment Scores:

Pred. No.:	4	53e+03	Length:	89934
Score:	92.50	Matches:	52	
Percent Similarity:	35.71%	Conservative:	38	
Best Local Similarity:	20.63%	Mismatches:	83	
Query Match:	8.66%	Indels:	79	
DB:	8	Gaps:	9	

US-10-024-955-7 (1-213) x AC007067 (1-89934)

```

/protein_id="CA20794.1"
/db_xref="GI:3581838"
/db_xref="SPTREMBL:O88043"
/translation="MSRTRMRLGLSVTALLTLTAALLSPSPAHEAVADYTTITVDPDSRGR
PAIDDIWGVFVEDINRAADGGIAELVYONSPFEYSTADNASTYPLTAMADGTAEVAV
NDGRLERNERNNTLSLAAGSTVNTGIRVGERGRDYSVMSRAGHGTTLVTLTL
DAAGTLAKAROVAKGOMRKYTATFTATPSNRRLAVTTTAPALDMVSLFPPDTYR
HOGSLKDLAEKTEALHPGLRPPGCIIVNTGSMEDYSADSGNQKRSYQMKTIIFR
VEERATIANFWGIVNOSTGLGTYETFRPAEDIGAPRLVFLVGVGGQNRKAVDDALL
KRHIDPLDLIEFANGSPRTSEWGKRAMEGRPFPHLTLEVGMEENLPKEFFARFEG
RYSLIEAKYDPDVTVNSGPDGSGATPDTAQLNRDRDGVNDDEHYNSPDMFLONDD
RYSYDNGRKPVELGEYASOGNAMKNALAEAFMTGERNADVKLASYAPLISNEDY
VOMSPEDLIFENNNHMSGSANLEYOKLFPMNNITGDSVSTATGTPSVSGITLGGVLS
WATSAADVDYKVTSSADGETLISDFDSQDSRWMAHSGMSWYODGOYVQDPAENNTM
VTADDPAMHDIYDLHVRATKSKGKGFVYAREGKDTGNTYKNNLGMNNTQSAFCQASD
GGKSTLKSAGSIEETGRAVIDIVKVRROYTLVLDGSEWGGEFTDQKPAEPRQVTVRDR
DRNGELIVKYVMADTARTAVDVGKAKVNSRAAVTTLAADDVAVNTGTADVPVTVS
TFSGAASEFTYTFPARNVTFILRIQR"
complement(5230..5233)
RBS
/note="possible RBS upstream of SC135.05c"
5336..5339
RBS
/note="possible RBS upstream of SC135.06"
5336..5339
/note="possible RBS upstream of SC135.06"
5361..5819
gene
/gene="SC135.06"
/note="SC01584"
5361..5819
CDS
/gene="SC135.06"
/note="SC135.06, unknown, len: 152 aa"
/codon_start=1
/translation="hypothetical protein SC135.06"
/protein_id="CA20795.1"
/db_xref="GI:3581839"
/db_xref="SPTREMBL:O88044"
/translation="MEMGELMPYDPDVLATYLAGHMRTRASVYDLAGLEGREGRTSPD
ALDGGGLIGRESGAFWQGVTRPAERATLTPRGSGPRHVFADGRPHGLDLSGSH

```

QY	136	GlmspphevalvaAlalaLeuSerleugCluileSerasp	148
Db	17700	CACGAC--CGCGNACGGCGGGCGGCCGCCACCGCGTCCACACGCTCTCGAGGCGATC	17644
Db	17643	GAGGACATCTTCGTCGCATGGCGTACGAGCTCGCGGAG	17605
RESULT	27		
LOCUS	AC007067		
DEFINITION	AC007067	Genomic sequence for Arabidopsis thaliana BAC T10024 from	linear PLN 28-JUN-2000
ACCESSION	AC007067	Chromosome 1, complete sequence.	
VERSION	AC007067.4	GI:4558521	
KEYWORDS	HTG.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE	Arabidopsis thaliana.		
AUTHORS	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 89934)		
AUTHORS	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Khan, S., Kim, C., Walker, M., Altafi, H., Araujo, R., Conn, L., Conway, A. B., Gonzalez, A., Hansen, N. F., Huizar, L., Kremetskaia, I., Lenz, C., Li, J., Liu, S., Luros, S., Rowley, D., Schwartz, J., Toriumi, M., Vysotskaia, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and Ecker, J. R.		
TITLE	Genomic sequence for Arabidopsis thaliana BAC T10024 from		
JOURNAL	Chromosome 1		
REFERENCE	2 (bases 1 to 89934)		
AUTHORS	Ecker, J. R.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA		
REFERENCE	3 (bases 1 to 89934)		
AUTHORS	Ecker, J. R.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-APR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA		
REFERENCE	4 (bases 1 to 89934)		
AUTHORS	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Khan, S., Kim, C., Walker, M., Altafi, H., Araujo, R., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Huizar, L., Kremetskaia, I., Lenz, C., Li, J., Liu, S., Luros, S., Rowley, D., Schwartz, J., Toriumi, M., Vysotskaia, V., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-JUN-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA		
REFERENCE	5 (bases 1 to 89934)		
AUTHORS	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Cheuk, R., Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bel, B., Chin, C., Chou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, J., Liu, S., Mukharshi, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thayer, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA		
COMMENT	On Apr 2, 1999 this sequence version replaced gi:4464282.		
FEATURES	location/Qualifiers		
SOURCE	1. 89934		
CDS	/organism="Arabidopsis thaliana"		
	/db_xref="taxon:3702"		
	/chromosome="1"		
	complement(120..350)		
	/note="unknown: similar to EST gb N96021"		
	/codon_start=1		
	/evidence=not_experimental		
	/product="T10024.1"		

RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-ORSAV;
 RA Hellig R.;
 RT "Proteococcus abyssii genome sequence: insights into archaeal chromosome structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ248284; CAB49318.1; -
 DR HSSP: O51739; 1AOR
 DR InterPro: IPR001203; Oxred_Ald_Fedxn.
 DR Pfam: PF01314; AFOR_C; 1.
 DR Pfam: PF02730; AFOR_N; 1.
 DR Complete proteome.
 SQ SEQUENCE 628 AA; 69690 MW; 264E6BB2A263D95 CRC64;

Query Match 8.8%; Score 93.5; DB 17; Length 628;
 Best Local Similarity 23.4%; Pred. No. 23;
 Matches 60; Conservative 35; Mismatches 74; Indels 87; Gaps 14;

QY 2 MKFLIAAVFAVAVSADPIHYDKITEINKAIDDAIAIEGSETIDPMKVPDHDHDKFER- 60
 DB 198 LKATAVAGDGEVY-ARPEFEKLEVEE-----LETAHEEYMSRR 238
 QY 61 -----HVGI-----VDF-----KGE-LAMR-NIEARG-----Lk 82
 DB 239 IMGTSRLLAANRIGVLPGRHFEPEVDYALVSGERLAEYNVKTGCFSCVVPSCSRVF 298
 QY 83 QMKROGQANVGE-----EGYKKAHLICGVH-DIVSMEDLAYKLG-DLHPTTHVISDIO 136
 DB 299 LIRKGYAGIMGEPEYALAGMTVRIGNSDLAALYAVKILNDGLDVISITSEVSWL- 357
 QY 137 DEVALLEISDEGNITMTSEFROFANVNHIGLSILDPFGVLSVLAIFODIVRK 196
 DB 358 -----MELYERDITS-----DEIGGLK---PVMQDMETVLTLDIADIARR 395
 QY 197 EMKVLAPEERLEK 212
 DB 396 GVGDLVADGVLAKEK 411

RESULT 7
 Q9XJUS PRELIMINARY; PRT; 697 AA.
 AC Q9XJUS;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE T10024.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBI_TaxID=3702.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Brooks S., Buehler E., Chao Q., Dunn P., Khan S., Kim C.,
 RA Walker M., Altati H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
 RA Hansen N.F., Huitzer L., Kremetskaia I., Lenz C., Li J., Liu S.,
 RA Lueros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
 RA Davis R.W., Federpspiel N.A., Theologis A., Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC T10024 from Chromosome 1."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007067; AAD39579.1; -
 DR InterPro: IPR000225; Armadillo.
 DR InterPro: IPR003613; ZnF_modrING.
 DR InterPro: IPR001841; ZnF_rING.
 DR Pfam: PF00514; Armadillo_seg; 4.
 DR SMART: SM00184; RING; 1.
 DR SMART: SM00504; Ubox; 1.
 DR PROSITE: PSS0176; ARM_REPEAT; 1.
 SQ SEQUENCE 697 AA; 76994 MW; F01AFEA2FC256F09 CRC64;

Query Match 8.7%; Score 92.5; DB 10; Length 697;
 Best Local Similarity 20.6%; Pred. No. 32;
 Matches 52; Conservative 38; Mismatches 83; Indels 79; Gaps 9;

QY 14 AVSADPIHYDKITEINKAIDDAIAIE---QSETIDPMKV-----P 52
 DB 139 SLSAFPAVSADLTTEVNEIDLIVVROARKYGVQGETNDKRAVSSINRILALFVNRVDP 198
 QY 53 DHAKFERHVI-----VDFKGLAMRNITARGLKOKKROGDANVKEEGVAKHL 103
 DB 199 DEINRILDHVGRKMGDCVKEINFLGE---EIDAEERIDERRK-----KSSDVEILL 248
 QY 104 LGV-----HDDIVSMEDLAYKLGDLHPTTHVISDIOFVVALS 143
 DB 249 LMGFICYRCILIGRIENDHNNHEDGIRKDHDLRL-----KVEDLCYIS 297
 QY 144 LEISDEGNITMTSEFROFANVNHIGLSILDPFGVLSVLAIFODIVRKENTK--- 200
 DB 298 LEIMTDPVVIEGTGYDR-SSITRWFGSGNITCPITGKITSTELVDNVSVROYIRKHCK 356
 QY 201 -----VLAPEFR 208
 DB 357 TNGIVLAGISR 368

RESULT 8

Q98PK8 PRELIMINARY; PRT; 979 AA.

AC Q98PK8;
 DT 01-OCT-2001 (TREMblrel. 18, Created)
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE P115-like (Mycoplasma hyorhinis) ABC transporter ATP-binding protein.
 DE MIPU_7140.
 GN MIPU_7140.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 NC NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=UAB CTIP;
 RX MEDLINE=21267165; Pubmed=11353084;
 RA Chambaud I., Hellig R., Ferris S., Barde V., Samson D., Galisson F.,
 RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis."
 RT Nucleic Acids Res. 29:2145-2153(2001).
 RL EMBL: AL445565; CAC13887.1; -
 DR Mypulist: MYPU_7140; -
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR003403; SMC_C.
 DR InterPro: IPR003395; SMC_N.
 DR Pfam: PF02483; SMC_C; 1.
 DR Pfam: PF02463; SMC_N; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 979 AA; 110333 MW; 4BC5C4BE1E54ACE CRC64;

Query Match 8.6%; Score 91.5; DB 16; Length 979;
 Best Local Similarity 21.2%; Pred. No. 61;
 Matches 46; Conservative 45; Mismatches 93; Indels 33; Gaps 9;

QY 3 KFLIAAVFAVAVSADPIHYDKIT-----EINKAIDDAIAIEQSETIDPMK 50
 DB 616 EFSVEAKVKEKLSNDKSPSLTAEVQTSIAKDLGESSSALDOQINEIOAKRVI--LQ 673
 QY 51 VPDHADFERNVIGVDFKGLAMRNITARGLKOKKROGDANVKEEGVAKHLIGVHD 109
 DB 674 AQIRAKK-----ESTLSLNTDPSKYVLEKNDEKRLRLNDSFSKITEKNAEFTLDNQ 729
 QY 110 DIVSMEDLAYKLG-DLHPTTHVISDIOFVVALSLEISDEGNITMTSEFROFANVNH 168

RESULT 6		
Q9Y1N1		
ID	Q9Y1N1	PRELIMINARY; PRT; 628 AA.
AC	Q9Y1N1;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)	
DE	Tungsten-containing aldehyde ferredoxin oxidoreductase (AOR-1).	
GN	PAB2085.	
OS	Pyrococcus abyssi.	
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;	
OC	Pyrococcus.	
OX	NCBI_TaxID=29292;	


```

RT "Three-dimensional structure of the immunophilin-like domain of
RT FKBP59 in solution."
RL Biochemistry 35:11045-11052(1996).
CC -1- FUNCTION: Component of unactivated mammalian steroid receptor
CC complexes that sediment at 8-10 S. May have a rotamase activity.
CC May play a role in the intracellular trafficking of hetero-
CC oligomeric forms of steroid hormone receptors.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBUNIT: Associates with HSP90 and HSP70 in unactivated steroid
CC hormone receptor complexes. Also interacts with peroxisomal
CC phytyl-CoA alpha-hydroxylase (PHH) (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- PTM: Phosphorylation by CK2 results in loss of HSP90 binding
CC activity.
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 FKBP-TYPE PPIASE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M84474; AAA31438.1; -
DR EMBL: M84986; AAA31439.1; -
DR PIR: A42386; A42386.
DR PDB: 1ROT; 07-DEC-96.
DR PDB: 1ROF; 07-DEC-96.
DR InterPro: IPR001179; FKBP_PPIase.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00254; FKBP. 2.
DR Pfam: PF00515; TPR. 3.
DR PROSITE: PS00453; FKBP_PPIASE. 1.
DR PROSITE: PS00454; FKBP_PPIASE. 2.
DR PROSITE: PS50059; FKBP_PPIASE. 3. 2.
DR Isomerase: Rotamase; TPR repeat; Repeat; Nuclear protein;
KM 3D-structure: Phosphorylation.
FT INIT MET 0
FT DOMAIN 49 137 PPIASE, FKBP-TYPE 1.
FT DOMAIN 166 252 PPIASE, FKBP-TYPE 2.
FT REPEAT 269 302 TPR 1.
FT REPEAT 318 351 TPR 2.
FT REPEAT 353 385 TPR 3.
FT MOD_RES 142 142 PHOSPHORYLATION (BY CK2).
FT CONFLICT 14 14 S->H (IN REF. 3).
FT CONFLICT 20 21 EG->T (IN REF. 3).
FT CONFLICT 25 25 S->T (IN REF. 3).
SQ SEQUENCE 457 AA; 51344 MW; ECS8CC4BCF66A44A CRC64;

Query Match 7.48; Score 79.5; DB 1; Length 457;
Best Local Similarity 22.18; Pred. No. 33;
Matches 27; Conservative 29; Mismatches 49; Indels 17; Gaps 3;

QY 18 DPVHYDKITEINKAIDAIAIPOSETI-----DKMVPDHD-KFERHY 62
DB 74 DKPSFDGKEGVKAMDAIAATKMGVLCRTCKPEYAYSAGSPPIIPNATVFE--V 131
QY 63 GIYDFKGLAMRNIEARGLKQMRQGDANVKGEGIVKAHLIGVHDDIVSMEDYLA 122
DB 132 ELFEFKGEDLTDEDDGGLIRIRTRTREGVARNPGALVEVALLEGYVYDRFDQELNFEV 191
QY 123 GD 124
DB 192 GE 193

RESULT 49
CH60_ERMCA STANDARD; PRT; 540 AA.

```

```

AC O66220;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOFA OR GROEL.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM 12633;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB008152; BA25237.1; -
DR HSP: P06139; 1GRL.
DR InterPro: IPR001844; Chaperin_Cpn60.
DR InterPro: IPR002423; Cpn60_TCP-1.
DR Pfam: PF00118; Cpn60_TCP1. 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60. 1.
KM Chaperone; ATP-binding.
KW NON_TER 540
SQ SEQUENCE 540 AA; 56292 MW; 7E41FEA699566719 CRC64;

Query Match 7.48; Score 79.5; DB 1; Length 540;
Best Local Similarity 18.68; Pred. No. 40;
Matches 44; Conservative 42; Mismatches 82; Indels 69; Gaps 9;

QY 13 VASADPIHYDKITEINKAIDA-IAIPOSETIDPMKVPDHDKFERHVGIVDFKGL 71
DB 107 VAAGNMNM-----DKRGIDKAVIAAVE--LKLALSPCSKSLAIOVGITANSNSD 156
QY 72 AMRNIEARGLKQMRQGDANVKGEGIVKAHLIGVHDD--IVSMEDYLA----- 120
DB 157 TVGKMAIEAMDKV-----KKEGVITVEEGGLDDELDVEGKQFDAGYISPYFINK 207
QY 121 -----KLGDLHPTHVYSIDIDPFVAALSLSEISDGSNTMTSFEV 159
DB 208 PEGVALESPFILLADKKTISNIREMLPEVAVAKAKPLVIAVEDVEGALATL----- 262
QY 160 RQFANVNHIGL-----SLDPIFGVSLDVLTAFODTVKREMTKVLAPAFKRELEK 212
DB 263 -----VYNTKRGIVKAAVAPGFC---DKRKMLQDIATLTGTGTVISEETGLELEK 311

RESULT 50
CH60_SALTY STANDARD; PRT; 547 AA.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

```

FT DOMAIN 1229 1397 LAMININ G-LIKE 3.
FT DOMAIN 1462 1633 LAMININ G-LIKE 4.
FT DOMAIN 1640 1813 LAMININ G-LIKE 5.
FT DOMAIN 431 523 COILED COIL (POTENTIAL).
FT DOMAIN 556 604 COILED COIL (POTENTIAL).
FT DOMAIN 655 717 COILED COIL (POTENTIAL).
FT DOMAIN 770 799 COILED COIL (POTENTIAL).
FT SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 82 91 BY SIMILARITY.
FT DISULFID 84 98 BY SIMILARITY.
FT DISULFID 101 110 BY SIMILARITY.
FT DISULFID 113 129 BY SIMILARITY.
FT DISULFID 132 146 BY SIMILARITY.
FT DISULFID 134 155 BY SIMILARITY.
FT DISULFID 157 166 BY SIMILARITY.
FT DISULFID 169 184 BY SIMILARITY.
FT DISULFID 187 202 BY SIMILARITY.
FT DISULFID 189 209 BY SIMILARITY.
FT DISULFID 212 221 BY SIMILARITY.
FT DISULFID 224 238 BY SIMILARITY.
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 269 269 INTERCHAIN (PROBABLE).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1088 1088 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1283 1283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1361 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 6 6 C -> S (IN REF. 2).
FT CONFLICT 18 18 C -> Y (IN REF. 2).
FT CONFLICT 248 248 G -> R (IN REF. 3).
FT CONFLICT 297 297 G -> A (IN REF. 3).
FT CONFLICT 431 433 THR -> HPS (IN REF. 2).
FT CONFLICT 679 679 S -> C (IN REF. 3).
FT CONFLICT 703 703 D -> G (IN REF. 2).
FT CONFLICT 706 706 N -> H (IN REF. 2).
FT CONFLICT 728 728 K -> R (IN REF. 2).
FT CONFLICT 730 730 F -> I (IN REF. 2).
FT CONFLICT 779 779 R -> G (IN REF. 1).
FT CONFLICT 810 810 R -> S (IN REF. 3).
FT CONFLICT 867 867 AEP -> QT (IN REF. 2).
FT CONFLICT 936 936 K -> E (IN REF. 3).
FT CONFLICT 970 970 L -> V (IN REF. 3).
FT CONFLICT 1132 1132 H -> R (IN REF. 2).
FT CONFLICT 1200 1200 F -> I (IN REF. 2).
FT CONFLICT 1382 1382 D -> A (IN REF. 2).
FT CONFLICT 1413 1414 NS -> EF (IN REF. 1).
FT CONFLICT 1489 1489 A -> S (IN REF. 2).
SQ SEQUENCE 1816 AA: 201818 MM: B49C45F3A4599DB CRC64;

```

Query Match 7.5%; Score 80; DB 1; Length 1816;
 Best Local Similarity 21.3%; Pred. No. 1.6e+02;
 Matches 44; Conservative 39; Mismatches 90; Indels 34; Gaps 9;

```

OY 25 ITEEINKAIDATAIAIEOSEIDPMKVPADHAEFERHVGIVDFEGLAMINIEARGKOM 84
DB 481 LQESINALDHVDAEDMNAI-TEKORDHEKHERRYKEQMEVYVAGALSADSALTIPOL 539
OY 85 K-ROGDANVKGEGIVAKHLLIGVHDDIVSEMYDLAYKLDELPHETH-VISDIODFVAL 142
DB 540 TLEELDEIRIKNASGI-----YAEIDGAKNELGKLSNLSLSDVLOEATDHYAINTL 590

```

```

OY 143 -----SLEISDEGNITMTSPEVRO-PANVYVHIG-----GLSTIDPIFGVLS 184
DB 591 QQADELSRNHSSDNGVQKALDASNVYENTANVSEANETRELALNTIDRIYAVSG 650
OY 185 VLTAIFODTVRKENTKVLAPAFKRELE 211
DB 651 IDFOIIVH--KDESDNLNOA--RELQ 673

RESULT 48
FKB4_RABIT
ID FKB4_RABIT STANDARD; PRT; 457 AA.
AC P27124;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE FK506-binding protein 4 (Possible pepitidyI-prolyl cis-trans isomerase
DE FKBp4) (EC 5.2.1.8) (Pliase) (p59 protein) (HSP binding
DE Immunophilin) (HBI) (FKBP52 protein) (52 kDa FK506 binding protein)
DE (FKBP59).
CN FKBp4 OR p59.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92165768; PubMed=1537818;
RA Lebeau M.-C., Massol N., Herrick J., Faber L.E., Renoit J.-M.,
RT "p59, an hsp 90-binding protein. Cloning and sequencing of its cDNA
RL and preparation of a peptide-directed polyclonal antibody."
RN J. Biol. Chem. 267:4281-4284(1992).
[2]
RP SEQUENCE OF 1-25.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=96154240; PubMed=8579355;
RA Deshpande K.L., Seubert P.H., Tillman D.M., Farkas W.R., Katze J.R.;
RT "Cloning and characterization of cDNA encoding the rabbit
RL tRNA-guanine transglycosylase 60-kilodalton subunit."
RN Arch. Biochem. Biophys. 326:1-7(1996).
[3]
RP DOMAINS.
RX MEDLINE=92335279; PubMed=1631118;
RA Callebaut I., Renoit J.-M., Lebeau M.-C., Massol N., Burny A.,
RA Baulieu E.-E., Morron J.-P.;
RT "An immunophilin that binds M(r) 90,000 heat shock protein: main
RL structural features of a mammalian p59 protein."
RN Proc. Natl. Acad. Sci. U.S.A. 89:6270-6274(1992).
[4]
RP PHOSPHORYLATION BY CK2.
RX MEDLINE=98070781; PubMed=9405642;
RA Miyata Y., Chambrud B., Radanyi C., Leclerc J., Lebeau M.-C.,
RA Renoit J.-M., Shirai R., Cateelli M.-G., Yahara I., Baulieu E.-E.;
RT "Phosphorylation of the immunosuppressant FK506-binding protein FKBp52
RL by casein kinase II: regulation of HSP90-binding activity of FKBp52."
RN Proc. Natl. Acad. Sci. U.S.A. 94:14500-14505(1997).
[5]
RP INTERACTION WITH PHYH.
RX MEDLINE=99162565; PubMed=10051602;
RA Chambrud B., Radanyi C., Camonis J.H., Rajkowski K., Schumacher M.,
RA Baulieu E.-E.;
RT "Immunophilins, refsum disease, and lupus nephritis: the peroxisomal
RT enzyme phytanoyl-CoA alpha-hydroxylase is a new FKBp-associated
RL protein."
RN Proc. Natl. Acad. Sci. U.S.A. 96:2104-2109(1999).
[6]
RP STRUCTURE BY NMR OF 1-148.
RX MEDLINE=96374215; PubMed=8780506;
RA Craescu C.T., Rouviere N., Popescu A., Cerpolini E., Lebeau M.-C.,
RA Baulieu E.-E., Mispelter J.;

```

DR Pfam: PF00313; CSD; 8.
 DR ProDom: PD000621; Cold_shock; 2.
 DR SMART: SM00357; CSP; 5.
 DR PROSITE: PS00352; COLD_SHOCK; 4.
 KW RNA-binding; Repeat.
 FT DOMAIN 26 87 CSD 1.
 FT DOMAIN 136 179 CSD 2 (INCOMPLETE).
 FT DOMAIN 186 245 CSD 3.
 FT DOMAIN 297 337 CSD 4 (INCOMPLETE).
 FT DOMAIN 349 410 CSD 5.
 FT DOMAIN 447 507 CSD 6.
 FT DOMAIN 519 579 CSD 7.
 FT DOMAIN 610 670 CSD 8.
 FT DOMAIN 674 735 CSD 9.
 SQ SEQUENCE 798 AA; 88894 MW; F484B3FA8B0955A4 CRC64;
 Query Match 7.5%; Score 80; DB 1; Length 798;
 Best Local Similarity 22.9%; Pred. No. 58;
 Matches 36; Conservative 29; Mismatches 56; Indels 36; Gaps 8;
 OY 15 VSADPIHYKITEEIKALIDDAIAIEQSETIDPMKVPPHAD-KPERHVGIY---DEKGE 70
 DB 574 VSAEYKVK---THSVNGITEEAPNTTYSKVIKRLPGVPTQIEYGMIEYEGDMKE 630
 OY 71 LAMRNIEAGLQMKRQGDANKYGEIGVKAHLILGVHDDIVSMEDLA----- 119
 DB 631 ---YVPEGIYGMANKGDLQKES--VKFQLCV-LGQNAQTMAYNITPLRAIVGVK 682
 OY 120 -----YKLGDLHPPTTHYISDIQDPYVALSLEISDE 149
 DB 683 DQGFPIYEVGDSKLLFVHKVQD---GIELQAGDE 716
 RESULT 47
 LMA4.MOUSE
 ID LMA4.MOUSE STANDARD: PRT: 1816 AA.
 AC P97927; P70409; 088785;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-4 chain precursor.
 GN LAMA4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 462-469; 478-483; 776-782 AND 940-945.
 RC STRAIN-BALB/c; TISSUE=Endothelial cells;
 RX MEDLINE=97363207; PubMed=9219532;
 RA Friesser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R., Sorokin L.M.;
 RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of endothelium."
 RL Eur. J. Biochem. 246:727-735(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=97202462; PubMed=9049981;
 RA Liu J., Mayne R.;
 RT "The complete cDNA coding sequence and tissue-specific expression of the mouse laminin alpha 4 chain."
 RL Matrix Biol. 15:433-437(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/c;
 RX MEDLINE=96010627; PubMed=9346933;
 RA Iivanainen A., Kortasmaa J., Sahberg C., Morita T., Bergmann U., Theiselff I., Tryggvason K.;
 RT "Primary structure, developmental expression, and immunolocalization of the murine laminin alpha4 chain."
 RL J. Biol. Chem. 272:27862-27868(1997).

RN [4]
 RP SEQUENCE OF 836-1106 FROM N.A.
 RC STRAIN=ICR; TISSUE=Placenta;
 RX MEDLINE=97296337; PubMed=9151674;
 RA Miner J.H., Patton B.L., Lentz S.I., Gilbert D.J., Snider W.D., Jenkins N.A., Copeland N.G., Sanes J.R.;
 RT "The laminin alpha chains: expression, developmental transitions, and chromosomal locations of alpha-5, identification of heterotrimeric laminins 8-11, and cloning of a novel alpha3 isoform."
 RL J. Cell Biol. 137:685-702(1997).
 RN [5]
 RP SEQUENCE OF 1467-1691 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=97187457; PubMed=9034910;
 RA Lentz S.I., Miner J.H., Sanes J.R., Snider W.D.;
 RT "Distribution of the ten known laminin chains in the pathways and targets of developing sensory axons."
 RL J. Comp. Neurol. 378:547-561(1997).
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN PERIPHERAL NERVES, CARDIAC MUSCLE, FAT, DERMIS, LUNG STROMA, AORTIC ENDOTHELIUM, ENDOCARDIUM AND ENDOTHELIUM OF BLOOD VESSELS IN SKIN AND BRAIN.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAIN G IS GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 3.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U58950; ABA41840.1; -
 DR EMBL: Y09827; CAA70970.1; -
 DR EMBL: U59865; AAC24725.1; -
 DR EMBL: U88352; AAC53178.1; -
 DR EMBL: U69176; AAC52982.1; -
 DR HSSP: P02468; IKT0.
 DR MGD: MGI:109321; Lama4.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR Pfam: PF00053; laminin_EGF; 3.
 DR Pfam: PF00054; laminin_G; 4.
 DR SMART: SM00181; EGF; 3.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 3.
 DR PROSITE: PS50025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 1816 POTENTIAL.
 FT DOMAIN 82 131 LAMININ ALPHA-4 CHAIN.
 FT DOMAIN 132 186 LAMININ EGF-LIKE 1.
 FT DOMAIN 187 240 LAMININ EGF-LIKE 2.
 FT DOMAIN 241 255 LAMININ EGF-LIKE 3.
 FT DOMAIN 256 825 LAMININ EGF-LIKE 4 (INCOMPLETE).
 FT DOMAIN 826 1030 DOMAIN II AND I.
 FT DOMAIN 1030 1222 LAMININ G-LIKE 1.
 FT DOMAIN 1042 1222 LAMININ G-LIKE 2.

```

RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT 'gene D' coding for an intracytoplasmic protein."
RL submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF150599; AAK30690.1; -
KW Antigen.
FT NON_TER 1 1
FT NON_TER 1010 1010
SQ SEQUENCE 1010 AA; 110694 MW; 785868DC92FF9C5B CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 1010;
Best Local Similarity 21.5%; Pred. No. 70;
Matches 53; Conservative 38; Mismatches 95; Indels 61; Gaps 11;

OY 5 LIIAAVAVYASADPIHYDK--ITEEI--NKAID---DAIAAIEQSETIDPKKVPDHADK 57
DB 401 LINTATALSGSMQDLINYNAGLTKEIDSNKQIDLIKEAATAILNNE-----KSDI 451
OY 58 PERHGVDFKELAMRNTEARGLKQKRGDANKGEGYV-----KAHL 104
DB 452 VEKQANIT---ALAEVTVNNKNLKP---DAKAVGAVALEIILKNDONTPLERSKML 502
OY 105 IGVHDDIVSM-----YDLAYKLGDLHPTTHVSDIODEFYVALSLEISDEG 150
DB 503 EATVAIVLSENLLEPKQKQMLEKAVDGLSLKDDASRAATIDGLKDYVTKSNLTEDG 562
OY 151 NITNM--SEFVKOPAVYNNHIGLSILDPIF--GYLSVLTALPQDVIKREKTKYLAPA 205
DB 563 TMLIVAGDKVNVSELSNMEKO---KLIGSVLAKGVEAGVLSPADQQLMQOHLMDKITAEQ 618

OY 206 FKRELEK 212
DB 619 TKKDTIK 625

RESULT 45
TPC2_CAEBL
ID TPC2_CAEBL STANDARD: PRT: 160 AA.
AC 009665;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Troponin C, isoform 2.
GN TNC-2 OR ZK673.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. TISSUE=Pharyngeal muscle;
RC STRAIN=Bristol N2;
RA Terami H., Kagawa H.;
RT "Functional characterization of the pharyngeal troponin C of
RT Caenorhabditis elegans.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- MISCELLANEOUS: THIS PROTEIN BINDS TWO CALCIUM IONS (POTENTIAL).
CC -1- TISSUE SPECIFICITY: Pharyngeal muscle.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB079299; BAB84566.1; -
DR EMBL; Z48585; CAA88482.1; -
DR HSSP; P02588; IPON.
DR WormPep; ZK673.7; CE01719.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 4.
DR Prodom; PD000012; EF-hand; 2.
DR SMART; SM00054; EFh; 4.
DR PROSITE; PS00018; EF_HAND; 2.
KW Calcium-binding; Repeat.
FT DOMAIN 28 39 ANCESTRAL CALCIUM SITE 1 (POTENTIAL).
FT CA_BIND 64 75 EF-HAND 2 (POTENTIAL).
FT DOMAIN 105 116 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT CA_BIND 141 152 EF-HAND 4 (POTENTIAL).
SQ SEQUENCE 160 AA; 18227 MW; 1852E2D8D5058CE CRC64;

Query Match 7.5%; Score 80; DB 1; Length 160;
Best Local Similarity 23.1%; Pred. No. 8.4;
Matches 27; Conservative 28; Mismatches 42; Indels 20; Gaps 3;

OY 53 DHADKFERHVGIVDFKGLAMR-----NIEARGLKQKRGDANKGE---- 95
DB 15 DQIEQFRKRYFMFDEKQGYRATQVQILRTMGAQFEERDKOLIKERDADGSGEIEFE 74
OY 96 --EGIVKAHLILGVHDDIVSMEDYLAAYKLGDLHPTTHV--ISDIQFVVALSLEISDE 149
DB 75 EFAAVAVAFVNNENDEGLEELRERAFRLYDEKGNGLVSDRLIIRALDDNVSEE 131

RESULT 46
UNR_RAT
ID UNR_RAT STANDARD: PRT: 798 AA.
AC P18395;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UNR protein.
GN UNR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=90370473; PubMed=2204029;
RA Jeffers M., Paclicucci R., Pellicer A.;
RT "Characterization of unr: a gene closely linked to N-ras.";
RL Nucleic Acids Res. 18:4891-4893(1990).
CC -1- FUNCTION: RNA-BINDING PROTEIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52311; CAA36549.1; -
DR PIR; S11210; S11210.
DR HSSP; P15277; IMJC.
DR InterPro; IPR002059; Cold_shock.

```



```

OS Erwinia aphidicola.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=68334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 14479;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB008153; BAA25239.1; -
DR HSSP: P06139; 1GRU.
DR InterPro: IPR001844; Chaperin_Cpn60.
DR InterPro: IPR002423; Cpn60_TCP-1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KM Chaperone; ATP-binding.
FT NON_TER 540 540
SQ SEQUENCE 540 AA; 56707 MW; B5CD54BBD39373E7 CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 540;
Best Local Similarity 17.7%; Pred. No. 33;
Matches 42; Conservative 46; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDA-IAAIEQSETIDPMKVPDADHFERHVGIVDEKGL 71
DB 107 VAAGMNPM-----DLKRIGIDQAVIAAVER-----LKLTVSCSDSKAIAQVGTISANSDA 156
QY 72 AMRNIERGLKQKROGDANVKEGEGIVKAHLIGVHDD---IVSMEDLAYLGLD--H 126
DB 157 TVGELLQAAMEKV-----GKEGVITVEEGTGIDDELVDVVEGMOFDRGYLSPYFINK 207
QY 121 -----KLGDLPTTHVISIDIDFVALSLTISDEGNITMTSFEV 159
DB 208 PETGAVELESPTLLADKRTISNIREMPLVEAVAKKPKPLIIIAEDVEGALATL----- 262
QY 160 KOFANVYNNHIGL-----SLIDPIFGVLSVLTALFODTVKEMTKVLAAPFKKLEK 212
DB 263 -----VYNTMRGIVKVAAYKAPGFG---DRRKAMLDIDIAVLGTGTVISSEIGMELEK 311

RESULT 43
CH60_ERWHE STANDARD; PRT; 541 AA.
ID CH60_ERWHE
AC 066216;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOEA OR GROEL.
OS Erwinia herbicola.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 7000;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB008150; BAA25233.1; -
DR HSSP: P06139; 1DK7.
DR InterPro: IPR001844; Chaperin_Cpn60.
DR InterPro: IPR002423; Cpn60_TCP-1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KM Chaperone; ATP-binding.
FT NON_TER 541 541
SQ SEQUENCE 541 AA; 56726 MW; A76172422E4559C5 CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 541;
Best Local Similarity 20.6%; Pred. No. 33;
Matches 42; Conservative 41; Mismatches 82; Indels 39; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDA-IAAIEQSETIDPMKVPDADHFERHVGIVDEKGL 71
DB 107 VAAGMNPM-----DLKRIGIDQAVIAAVER-----LKLTVSCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIERGLKQKROGDANVKEGEGIVKAHLIGVHDD---IVSMEDLAYLGLD--H 126
DB 157 TVGQLLAQAAMEKV-----GKEGVITVEEGTGIDDELVDVVEGMOFDRGYLSPYFINK 207
QY 127 PTHVISIDIDFVALSLTISDEGNITMTSFEVROFANVYNNH---GGLSLIDPIFGVLS 184
DB 208 PETGAVELESPTLLADKRTISN-----IREMPLVEAVAKKPKPLIIIAEDVEGE 257
QY 185 VLTAFIDQTVR--KEMTKVLAAPF 206
DB 258 ALATVYNTMRGIVKVAAYKAPGF 281

RESULT 44
SCA4_RICPA STANDARD; PRT; 1010 AA.
ID SCA4_RICPA
AC 09A175;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (Pst120) (120 kDa antigen)
DE (Protein ps 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia parkeri.
OC Bacteria: Proteobacteria: alpha subdivision: Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35792;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;

```

```

CH60_ENTAS
ID CH60_ENTAS STANDARD: PRT; 539 AA.
AC 06190;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROL OR MOBA OR GROEL.
OS Enterobacter asburiae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxId=61645;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-JCM 6051;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB008137; BAA25207.1; -
CC DR HSSP; P06139; LJON.
CC DR InterPro; IPR001844; Chaperin_Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCP-1.
CC DR Pfam; PF00118; Cpn60_TCP1.1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXCPI.
CC DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC FT NON_TER 539 539
CC SQ SEQUENCE 539 AA; 56399 MW; 8A8A5CAF9ADBDAFB CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 539;
Best Local Similarity 18.1%; Pred. No. 33;
Matches 43; Conservative 43; Mismatches 82; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDAI-ATAEQSETIDPMKVPDHADEFERHVGIVDEKGEI 71
DB 107 VAAGMNP-----DLKRIDKAVVAAYEE--LKALSPCDSKAIAQVGTISANSDE 156
QY 72 AMNINERAGLKOKKRGDANVKEGEGIVAKHLIGVHD---IVSEYDAY----- 120
DB 157 TVGKLLAEMDKV-----GKEGVITYEDGTGLEDELVDVEGQFDFGYSIPYFINK 207
QY 121 -----KLGDLHPTTHVISDIQDFVALSLEISDEGNTMTSFEV 159
DB 208 PETGAVELESFILLADKTKISIREMLPYLEAVAKKAGKPLVITAEVDEGALATL----- 262
QY 160 RQFANVYNIHIGL-----SILDPITFGLSDVLTALIFQDTPVAKEMTKYLAPAFKRELEK 212
DB 263 -----VVNTMRGIVKVAAYKAPGFG--DRKKMLQDIATLTGTVISEIGMELEK 311

```

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROL OR MOBA OR GROEL.
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxId=548;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-JCM 1235;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB008141; BAA25215.1; -
CC DR HSSP; P06139; LJON.
CC DR InterPro; IPR001844; Chaperin_Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCP-1.
CC DR Pfam; PF00118; Cpn60_TCP1.1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXCPI.
CC DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC FT NON_TER 540 540
CC SQ SEQUENCE 540 AA; 56342 MW; A55A23E3C21939B9 CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 540;
Best Local Similarity 18.1%; Pred. No. 33;
Matches 43; Conservative 43; Mismatches 82; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIEQSETIDPMKVPDHADEFERHVGIVDEKGEI 71
DB 107 VAAGMNP-----DLKRIDKAVVAAYEE--LKALSPCDSKAIAQVGTISANSDE 156
QY 72 AMNINERAGLKOKKRGDANVKEGEGIVAKHLIGVHD---IVSEYDAY----- 120
DB 157 TVGKLLAEMDKV-----GKEGVITYEDGTGLEDELVDVEGQFDFGYSIPYFINK 207
QY 121 -----KLGDLHPTTHVISDIQDFVALSLEISDEGNTMTSFEV 159
DB 208 PETGAVELESFILLADKTKISIREMLPYLEAVAKKAGKPLVITAEVDEGALATL----- 262
QY 160 RQFANVYNIHIGL-----SILDPITFGLSDVLTALIFQDTPVAKEMTKYLAPAFKRELEK 212
DB 263 -----VVNTMRGIVKVAAYKAPGFG--DRKKMLQDIATLTGTVISEIGMELEK 311

```

```

RESULT 41
CH60_ENTAE STANDARD: PRT; 540 AA.
AC 06198;
DT 30-MAY-2000 (Rel. 39, Created)

```

```

RESULT 42
CH60_ERWAP STANDARD: PRT; 540 AA.
AC 066222;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROL OR MOBA OR GROEL.

```

QY	105	IGV-----HDDIVSMEDLAKGDLHPTTHVHSIDODFVVALSLE 145
Db	121	IDILADEEHEDHMELETDLIARLG-----IONVAQAQITLLE 156
RESULT 38		
VATE_SULTO	STANDARD;	PRT: 194 AA.
AC	Q971B8;	
DT	15-JUN-2002 (Rel. 41, Created)	
DT	15-JUN-2002 (Rel. 41, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	V-type ATP synthase subunit E (EC 3.6.3.14) (V-type ATPase subunit E).	
NC	AFPE OR S11435.	
OS	Sulfolobus tokodaii.	
OC	Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae; Sulfolobus.	
CC	NCBI_TaxID=111955;	
CC	[1]	
CC	SEQUENCE FROM N.A.	
CC	STRAIN-JCM 10545 / 7;	
CC	MEDLINE=21456156; PubMed=11572479;	
CC	Kawarabayashi Y., Hino Y., Horikawa H., Jin K., Takahashi M., Sekine M., Baba S.-I., Ankei A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y., Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A., Oshima T., Kikuchi H.;	
CC	*Complete genome sequence of an aerobic thermacidophilic Crenarchaeon, Sulfolobus tokodaii strain7.*;	
CC	DNA Res. 8:123-140(2001).	
CC	-I- FUNCTION: Produces ATP from ADP in the presence of a proton gradient across the membrane.	
CC	-I- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate + H(+)(Out).	
CC	-I- SIMILARITY: BELONGS TO THE V-ATPASE E SUBUNIT FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AP000986; BAB6502.1; -	
DR	InterPro; IPR002842; ATPsynth_Esub.	
DR	Pfam; PF01991; V-ATP-synt_E.1.	
KW	Hydrolase; ATP synthetis; Hydrogen ion transport;	
KM	Complete proteome.	
SO	SEQUENCE 194 AA: 22634 MW; 20P46AF95783449A CXC64;	
QY	Query Match	7.5%; Score 80.5; DB 1; Length 194;
	Best Local Similarity	20.6%; Pred. No. 9.7;
	Matches 41; Conservative 39; Mismatches 68; Indels 51; Gaps 9;	
QY	22 YDKITEENKKAIDQAIAMIEQSETIDPPKAVDPHADKFERHNGIVFKEELAMRNIEAGL 81	
Db	26 FKIIISSENQIITIDVAVERYRS--AKITDLVKNNDRI---RGEIAMEIENKRL 77	
QY	82 KQKROGQANVKGEGICVAKHLILGVHDDIVSMEDVLAVKGLDHPHTTHVHSIDODFVVA 141	
Db	78 --ISEKMYWLENVKNKKSLIEVVKD-----NYKKG-----LEST----- 113	
QY	142 LSLSEISDEGNITWTSFEYRQEPANVNH-----IGGLSLDPIFEGLSD----- 184	
Db	114 ISREVSDESIIYCSFDQKSIISDIKKKKRISKIVADEKIVGIIYVPDSSLKDPITLE 173	
QY	185 -VLTAIFODVTKKEMTKVL 202	

```

Db 174 TILNOVFDD-IRDKIAQIL 191

RESULT 39
CH60_ENTAG STANDARD: PRT: 539 AA.
AC 066200:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOXA OR GROEL.
OS Enterobacter agglomerans (Pantoea agglomerans).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
RX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 1236;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RT J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1 FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1 SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1 SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/).license@isb-sib.ch).
CC -----
CC EMBL: AB008142; BAA25217.1; -.
CC HSSP: P06139; LJON.
DR InterPro: IPR001844; Chaprinin.Cpn60.
DR InterPro: IPR002423; Cpn60/TCF-1.
DR Pfam: PF00118; Cpn60_TCF1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCPL.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KM Chaperone; ATP-binding.
FT NON_TER 539 539
SQ SEQUENCE 539 AA; 56373 MW; 6583295F73299C3B CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 539;
Best Local Similarity 18.1%; Pred. No. 33;
Matches 43; Conservative 43; Mismatches 82; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKTEEINKAIDDATA-AIEQSETIDPKKVPDHDKFERHWGYDFKGL 71
DB 107 VAAQNPW-----DLRGIDKAVASAVEE---LKALSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIENAGLKQMKRQGANVKGEGGIYKAHLLGVHDD---IYSMEVDLAY----- 120
DB 157 TVGKLTIAEMKV-----GKGVTITVEDGTGLELDLVDVEGMQFDRGLSPYFLNK 207
QY 121 -----KLGDLHTTFVHYSIDIOFVALSLEIIDEQNTWTSTFEV 159
DB 208 PETGAVELESPILLADKKISIRIMLPVLEVAAGAKGRIVIIADVDGEALATL----- 262
QY 160 ROFANVNHIGL-----SILDPFGVSLVDVLTALFDQTVRKEMTKVLAIPARKELEK 212
DB 263 -----VVVTTMRGIYVAAVKAQGF-----DRRKAMLDIATLTGLGTIVISEIGMELEK 311

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:355-364(1998).
 [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC TISSUE=pituitary;
 RA Song H., Peng Y., Zhou J., Huang Q., Dai M., Mao Y., Xu Y., Xu X.,
 RA Luo M., Hu R., Chen J.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE OF 1-498 FROM N.A. (SHORT ISOFORM).
 RC TISSUE=Brain;
 RA Xu W., Gibbs R.A.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 [4]
 RP INTERACTION WITH UNRIP.
 RX MEDLINE=99158740; PubMed=10049359;
 RA Hunt S.L., Hsuan J.J., Totty N., Jackson R.J.;
 RT "unt, a cellular cytoplasmic RNA-binding protein with five cold-shock
 RT domains, is required for internal initiation of translation of human
 RT rhinovirus RNA."
 RL Genes Dev. 13:437-448(1999).
 CC -1- FUNCTION: RNA-BINDING PROTEIN. REQUIRED FOR INTERNAL INITIATION OF
 CC TRANSLATION OF HUMAN RHINOVIRUS RNA.
 CC -1- SUBUNIT: INTERACTS WITH UNRIP.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PROBABLY PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB020692; BAA74908.1; -
 DR EMBL: AF07054; AAD27787.1; -
 DR EMBL: AF070542; AAC28634.1; -
 DR HSSP: P15277; IMTC.
 DR MIM: 191510; -
 DR InterPro: IPR002059; Cold_shock.
 DR Pfam: PF00313; CSD; 8.
 DR ProDom: PD00621; Cold_shock; 2.
 DR SMART: SM00357; CSP; 5.
 DR PROSITE: PS00352; COLD_SHOCK; 4.
 KW RNA-binding; Repeat; Alternative splicing.
 FT DOMAIN 26 87 CSD 1.
 FT DOMAIN 136 179 CSD 2 (INCOMPLETE).
 FT DOMAIN 186 245 CSD 3.
 FT DOMAIN 297 337 CSD 4 (INCOMPLETE).
 FT DOMAIN 349 410 CSD 5.
 FT DOMAIN 447 507 CSD 6.
 FT DOMAIN 519 579 CSD 7.
 FT DOMAIN 610 670 CSD 8.
 FT DOMAIN 674 735 CSD 9.
 FT VARSPIC 104 134 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 31 31 E -> G (IN REF. 2).
 FT CONFLICT 92 95 OEIL -> TRNP (IN REF. 2).
 FT CONFLICT 243 243 T -> P (IN REF. 2).
 FT CONFLICT 363 385 IKCVDRVDMFHFSEILDGNL -> HPSVIRRCVCSF

FT SEQUENCE 798 AA: 88884 MW: 28003232F33D454DA CRC64;
 SQ
 Query Match 7.6%; Score 81; DB 1; Length 798;
 Best Local Similarity 22.9%; Pred. No. 48;
 Matches 36; Conservative 29; Mismatches 56; Indels 36; Gaps 8;
 OY 15 VSADPIHYKITEINKAIDDAIAIASETIDPMKVPDHAD-KERHVGIV---DFKGE 70
 DB 574 VSAEKYKN---THSVGITEADPTIYSGKVRPLRSVPPTQTEYGMIEYEDGMKE 630
 OY 71 LAMRNIENGKLOMKRQGANVKGEGIVKAHLICVHDIVSMEYDLA----- 119
 DB 631 ----VYPPGIVGMANKGGLQKGES--VKFOLCV-IGQNAQTMAVNIPLRRATVECVK 682
 OY 120 -----YKLGDLHPPTHYISDIODPVVALSLSEISDE 149
 DB 683 DQGFPIYEVGDSKLEFFHKEVOD---GIELQAGE 716
 RESULT 37
 BFR_SERMA
 ID BFR_SERMA STANDARD; PRT; 159 AA.
 AC 068935;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Bacterioferritin (BFR) (Cytochrome B-557).
 GN BFR.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Serratia.
 OX NCBI_TaxID=615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Noorani S.M., Lindahl L., Zengel J.M.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May perform analogous functions in iron detoxification
 CC and storage to that of animal ferritins (by similarity).
 CC -1- SUBUNIT: OLIGOMER OF 24 IDENTICAL SUBUNITS (BY SIMILARITY).
 CC -1- MISCELLANEOUS: Bacterioferritin contains protolome IX in addition
 CC to the non-hem iron core (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE BACTERIOFERRITIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF058451; AAC14293.1; -
 DR HSSP: P11056; IBCF.
 DR InterPro: IPR002024; Bacterioferritin.
 DR Pfam: PF01334; Bacteriofer; 1.
 DR PRINTS: PR00601; BACTERIFERTIN.
 DR ProDom: PD002269; Bacterioferritin; 1.
 DR TIGRfams: TIGR00754; bfr; 1.
 DR PROSITE: PS00549; BACTERIOFERRITIN; 1.
 KW Iron storage; Heme.
 FT METAL 52 52 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT SEQUENCE 159 AA: 18553 MW: 831A865EA98C82D CRC64;
 Query Match 7.5%; Score 80.5; DB 1; Length 159;
 Best Local Similarity 24.6%; Pred. No. 7.6;
 Matches 41; Conservative 19; Mismatches 46; Indels 61; Gaps 8;
 OY 24 KITEENKAIIDAIAIIEQ-----SETIDPMKVPDADKFERIV 62
 DB 6 KIIAHNKLILGNELVAINOYFLHARFKWGLRLNDKREYSIDEM--HADRIERI 62
 OY 63 ----GIVDFKGLAMRNI-----EARGLKQMKRQGANVKGEGIVKAHL 104

```

RT "Sequence analysis of the genome of the uncultured cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-116(1996).
CC -I- FUNCTION: MAY PLAY A ROLE IN THE OSMOTIC ADAPTATION (BY
CC SIMILARITY).
CC -I- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (EI-E2 ATPASES). SUBFAMILY IB.
CC -I- SIMILARITY: CONTAINS 1 HMA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90904; BAA17268.1; -.
DR HSSP: P04129; IAF1.
DR InterPro: IPR00157; ATPase_EI-E2.
DR InterPro: IPR00154; Heavy_membrane_transp.
DR Pfam: PF00142; EI-E2_ATPase; 1.
DR Pfam: PF00403; HMA; 1.
DR Pfam: PF00702; Hydrolyase; 1.
DR PROSITE: PS00154; ATPase_EI-E2; 1.
DR PROSITE: PS01047; HMA; 1.
DR PROSITE: PS00846; HMA_2; 1.
KM Hydrolyase: Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KM Metal-binding; Complete proteome.
FT DOMAIN 1 94 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 95 115 POTENTIAL.
FT DOMAIN 116 125 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 126 145 POTENTIAL.
FT DOMAIN 146 152 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 153 173 POTENTIAL.
FT DOMAIN 174 193 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 194 214 POTENTIAL.
FT DOMAIN 215 342 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 343 365 POTENTIAL.
FT DOMAIN 366 372 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 373 390 POTENTIAL.
FT TRANSMEM 391 543 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 544 564 POTENTIAL.
FT DOMAIN 565 687 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 688 707 POTENTIAL.
FT TRANSMEM 708 719 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 720 738 POTENTIAL.
FT DOMAIN 739 745 EXTRACELLULAR (POTENTIAL).
FT MOD_RES 4 69 HMA.
FT MOD_RES 428 428 PHOSPHORYLATION (PROBABLE).
FT METAL 14 14 POTENTIAL.
FT METAL 17 17 POTENTIAL.
FT METAL 633 633 MAGNESIUM (BY SIMILARITY).
FT METAL 637 637 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 745 AA; 79950 MW; 40831F66621735FB CRC64;

Query Match 7.6%; Score 81.5; DB 1; Length 745;
Best Local Similarity 25.2%; Pred. NO. 40;
Matches 30; Conservative 26; Mismatches 46; Indels 17; Gaps 5;

```

```

RESULT 35
RS4E_METVA STANDARD; PRT; 244 AA.
ID RS4E_METVA
AC P14023;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S4e.
GN RPS4E.
OS Methanococcus vannielii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2187;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=90040717; PubMed=2530355;
RA Auer J., Spicker G., Boeck A.;
RT "Organization and structure of the Methanococcus transcriptional unit
RT homologous to the Escherichia coli 'spectinomycin operon'.
RT Implications for the evolutionary relationship of 70 S and 80 S
RT ribosomes."
RT J. Mol. Biol. 209:21-36(1989).
RL -----
CC -I- SIMILARITY: BELONGS TO THE S4E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16720; CAA34692.1; -.
DR PIR: S05616; RMX4.
DR InterPro: IPR000302; KOW_motif.
DR InterPro: IPR000876; Ribosomal_S4E.
DR InterPro: IPR002942; S4.
DR Pfam: PF00467; KOW; 1.
DR Pfam: PF00900; Ribosomal_S4e; 1.
DR Pfam: PF01479; S4; 1.
DR PRODOM: PD002667; Ribosomal_S4E; 1.
DR SMART: SM00363; S4; 1.
DR PROSITE: PS00528; RIBOSOMAL_S4E; 1.
KM Ribosomal protein.
SQ SEQUENCE 244 AA; 27325 MW; 754FC1D9784EF5B5 CRC64;

Query Match 7.6%; Score 81; DB 1; Length 244;
Best Local Similarity 22.4%; Pred. NO. 12;
Matches 36; Conservative 29; Mismatches 72; Indels 24; Gaps 6;

```

```

DB 17 ADPIHDKTEETENKAIDAIAIROSETIDPMKVPDHDKFERHVGIVDFKGLAMRNI 76
DB 56 ADNAREAKTIOMGKILIDVKKRKEKHPVGLMDVLSPELNEVNIYLVDEGRISLKT 115
QY 77 EARGIKOMKROGDANVKGEGIVKAHLIGVHDIVSMEYDLAVYGLDHPHTHYI 132
DB 614 KAOQVEQLOOKKNIVAMVDGINDAPALQAQADVGI--AIGCTDVAIAASDI--TLTA 667
QY 133 SDIOFVALSLEISDEGNITMTSEFVROFANVNHIG--GLSIDPIFGVLSVLT 188
DB 668 GLOGLITRAIKSRATMGIRNLF----FAPIYVIGIPVAAGLFYPLGILLNITLA 722

RESULT 36
UNR_HUMAN STANDARD; PRT; 798 AA.
ID UNR_HUMAN
AC Q75534; Q94961; Q9Y254;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UNR protein.
DE UNR OR D15135E OR KIAA0885.
OS Homo sapiens (Human).

```

CC Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 6986;
RA Harada H., Ishikawa H.;
RT "Phylogenetic relationship based on groE genes among phenotypically
related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwina
species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB008151; BAA25235.1; -.
DR HSSP; P06139; 1DK7.
DR InterPro; IPR001844; Chaprinin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
FT NON_TER 541 541
SQ SEQUENCE 541 AA; 56817 MW; 7111AA03A703E543 CRC64;

Query Match 7.6%; Score 81.5; DB 1; Length 541;
Best Local Similarity 18.1%; Pred. No. 28;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

OY 13 VAVSADPIHYDKITEINKAIDA-IAAIEOSETIDPMKVPDHADKFERHVGIVDFKGL 71
DB 107 VAAGNMPPM-----DLKRGIDQAVIAAVE--LKLKSLVPCSDSKAIAOVGTISANSDE 156
OY 72 AMRNTAARGLKQKRGDANVKGEGIVKAHLIGVHD--IYAMEYDLAY----- 120
DB 157 TVGTILIAQAMEKV-----GKEGVITVEGTGLDGLDVEGMOQDFRGILSPYFINK 207
OY 121 -----KLGLDLPHTTHVTSIDIDFVALSLDISDEGNITMTSEFV 159
DB 208 PETGAVELETPTILLADKISINIREMLPVLEAVAKAGKPLLIADVEGEALATL----- 262
OY 160 ROFANVNHIGL-----SILDPIFGVLSVLTALFODTVRKEMTKVLAFAFKRELEK 212
DB 263 -----VVNTMRGIVKAAVAKAPFG--DRKRAMLDIALITLGTIVISBEIGMELEK 311

RESULT 33
CH60_HAEDU
ID CH60_HAEDU STANDARD; PRT; 547 AA.
AC P31294;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (protein Cpn60) (groEL protein).
GN GROEL OR MOPA OR GROEL.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=93014108; PubMed=1356926;
RA Parsons L.M., Waring A.L., Shayegani M.;
RT "Molecular analysis of the Haemophilus ducreyi groE heat shock
operon.";
RL Infect. Immun. 60:4111-4118(1992).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M91030; AAA24961.1; -.
DR PIR; B49203; B49203.
DR HSSP; P06139; 1GRL.
DR InterPro; IPR001844; Chaprinin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 547 AA; 57812 MW; 3CAF933AD98B45E CRC64;

Query Match 7.6%; Score 81.5; DB 1; Length 547;
Best Local Similarity 18.2%; Pred. No. 28;
Matches 43; Conservative 44; Mismatches 82; Indels 67; Gaps 9;

OY 13 VAVSADPIHYDKITEINKAIDAIAIEOSETIDPMKVPDHADKFERHVGIVDFKGL 72
DB 107 VAAGNMPPM-----DLKRGIDQAVIAAVE--VELKAISKPCPTSEIRQVGTISANSDET 157
OY 73 MRNTAARGLKQKRGDANVKGEGIVKAHLIGVHD--DIV-SNEYDLAY----- 120
DB 158 VGKILIAQAMEKV-----GKEGVITVEGTGLDGLDVEGMOQDFRGILSPYFINKP 208
OY 121 -----KLGLDLPHTTHVTSIDIDFVALSLDISDEGNITMTSEFVR 160
DB 209 EAGTYELDNPIYLLVDKISINIREMLPVLEAVAKAGKPLLIADVEGEALATL----- 262
OY 161 OFANVNHIGL-----SILDPIFGVLSVLTALFODTVRKEMTKVLAFAFKRELEK 212
DB 263 -----VVNTMRGIVKAAVAKAPFG--DRKRAMLDIALITLGTIVISBEIGMELEK 311

RESULT 34
ATCS_SYNY3
ID ATCS_SYNY3 STANDARD; PRT; 745 AA.
AC P73241;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cation-transporting ATPase pacs (EC 3.6.3.-).
GN PACS OR SLL1920.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,
RA Miyaajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraaki A., Nakazaki N., Nario K.,
RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;

SQ SEQUENCE 461 AA; 49166 MW; CADD4483E758FEAD CRC64;
 Query Match 7.7%; Score 82; DB 1; Length 461;
 Best Local Similarity 20.4%; Pred. No. 21;
 Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;
 QY 23 DKITEE-----INKAIDAIATIEOSETIDPMKVPDHDHAKFE-----RH 61
 DB 219 NMFRTKGIHLTKA-----SISATIEESONQVRIYNDQVEEDYVAIGROFNASTIGLDN 275
 QY 62 VGIV-DFKGLANRNIANGKMKRQGD-----ANKGEEGIVKAHLIGVHDDIYS 113
 DB 276 AGVIRDRGVIPVDETMRTPNPIVAIGDITGKMLAHVASHOGVIAAKNISGHNH-----V 332
 QY 114 MEVDLAVKIGDLHPHTHTVSDIQDFVALSLETSDEGNT--TMTSEVHQ----- 161
 DB 333 MDLSAIPSVIETHP-----EIAWGLSLQEAEOQNPAAKTKRPFKAIGKAVAGAS 384
 QY 162 --FANVNNH-----IGLSILDPFCVLSDLVAIFQDVRKREMT 199
 DB 385 DGFMAIVSHETIQIIGAVIVGPHASSLIGEMTL-----AIRNBLT 425

RESULT 31
 MYS3_SCHPO STANDARD: PRT: 2104 AA.
 ID MYS3_SCHPO 014157: 042730:
 AC 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE MYO3 in type II heavy chain 2.
 GN MYO3 OR MYP2 OR SPAC4A8.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98119390; PubMed=9459302;
 RA Motegi F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;
 RT Identification of MYO3, a second type-II myosin heavy chain in the
 RT fission yeast Schizosaccharomyces pombe.*;
 RL FEBS Lett. 420:161-166(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98062346; PubMed=9398685;
 RA Bezanilla M., Forsburg S.L., Pollard T.D.;
 RT Identification of a second myosin-II in Schizosaccharomyces pombe;
 RT Myo2p is conditionally required for cytokinesis.*;
 RL Mol. Biol. Cell 8:2693-2705(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Galliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Brooks J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
 RA Gholizadeh S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor R., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Pitz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Carrutell L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.*";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: STABILIZES THE F-ACTIN CABLES FORMING THE F-ACTIN RING
 CC THAT SURROUNDS THE NUCLEUS DURING INTERPHASE. MAY WORK IN
 CC CONJUNCTION WITH MYO2.
 CC -1- SUBUNIT: BINDS TO CDC4 AND RLC1.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).

DR EMBL: AB007633; BAA24579.1; -
 DR EMBL: AF029788; AAC04615.1; -
 DR EMBL: 298762; CAB11475.1; -
 DR HSPD: P08799; 1MND.
 DR Interpro: IPR004009; Myosin_N.
 DR Interpro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_N.
 DR Pfam: PF02736; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR Prodom: PD000355; myosin_head; 1.
 DR SMART: SM00242; MYSC; 1.
 DR MYO3in: collid coll: ATP-binding; Actin-binding; Alkylation.
 KW MYO3in: collid coll: ATP-binding; Actin-binding; Alkylation.
 FT DOMAIN 1 828 MYOSIN HEAD-LIKE.
 FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
 FT DOMAIN 646 660 ATP-BINDING (POTENTIAL).
 FT NP_BIND 182 189 ATP (POTENTIAL).
 FT MOD_RES 686 686 ALKYLATION (BY SIMILARITY).
 FT CONFLICT 1193 1193 D -> G (IN REF. 2).
 FT CONFLICT 1304 1304 E -> G (IN REF. 2).
 FT CONFLICT 1344 1344 E -> K (IN REF. 2).
 FT CONFLICT 1420 1420 G -> D (IN REF. 2).
 SQ SEQUENCE 2104 AA; 242570 MW; 3A0548594028D258 CRC64;
 Query Match 7.7%; Score 82; DB 1; Length 2104;
 Best Local Similarity 24.8%; Pred. No. 13e+02;
 Matches 36; Conservative 25; Mismatches 34; Indels 30; Gaps 6;
 QY 24 KITEINKAIDAIAIEOSETIDPMKVPDHDHAKFEHVGIVDFKGLAMRN----- 75
 DB 1864 EVSEQJLKKKFFDSSARPF-----DEIEMEKEVLTLKSNLAKQKDLSSLSIVE 1908
 QY 76 -IEARGLKMKRQGDANVKGEGIVKAHLIGVHDDIYSMEVDLAVKIGDLHPHTHTVSD 134
 DB 1909 RIQIEMFALTKQKDSNNHNEML-QLHROLGV---IQREKDLDELKLPDLTKTYVID 1964
 QY 135 IODFVALSLETSID-EGNITMTSE 158
 DB 1965 SKD-VRMLOKQISDLSPASDIE 1988
 RESULT 32
 CH60_PANAN STANDARD: PRT: 541 AA.
 ID CH60_PANAN 066218:
 AC 066218;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
 GN GROEL OR MORA OR GROEL.
 OS Pantoea ananas (Erythrina uredoformans).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

DR InterPro: IPR000089; Biotin_lipoyl.
DR InterPro: IPR000901; CPSase.
DR InterPro: IPR000022; Carboxyl_trans.
DR Pfam: PF00289; CPSase_L_chain: 1.
DR Pfam: PF00364; Biotin_lipoyl; 1.
DR Pfam: PF02785; Biotin_carb_C; 1.
DR Pfam: PF02786; CPSase_L_D2; 1.
DR PROSITE: PS00188; BIOTIN; 1.
DR PROSITE: PS00866; CPSASE_1; 1.
DR PROSITE: PS00867; CPSASE_2; 1.
KM Fatty acid biosynthesis; Biotin; Ligase; Multifunctional enzyme;
KM ATP-binding; Phosphorylation; Alternative splicing.
FT NP_BIND 458 463 ATP (POTENTIAL).
FT ACT_SITE 584 584 BIOTIN.
FT BINDING 927 927 BIOTIN.
FT DOMAIN 2095 2124 COENZYME A-BINDING (POTENTIAL).
FT MOD_RES 219 219 PHOSPHORYLATION (BY SIMILARITY).
FT VARSPIC 1115 1215 MISSING (IN SHORT ISOFORM).
FT CONFLICT 1365 1365 S -> N (IN REF. 2).
FT CONFLICT 1375 1375 T -> S (IN REF. 2).
FT CONFLICT 1408 1408 D -> G (IN REF. 2).
FT CONFLICT 1421 1421 P -> T (IN REF. 2).
FT CONFLICT 1450 1450 V -> M (IN REF. 2).
FT CONFLICT 1487 1487 P -> T (IN REF. 2).
FT CONFLICT 1555 1558 EGRY -> KEGV (IN REF. 2).
FT CONFLICT 1609 1609 N -> S (IN REF. 2).
FT CONFLICT 1630 1630 N -> F (IN REF. 2).
FT CONFLICT 1844 1844 A -> P (IN REF. 2).
FT CONFLICT 1917 1918 M1 -> IM (IN REF. 2).
SQ SEQUENCE 2483 AA; 279690 MW; A8C678276A9F753 CRC64;

Query Match
Best Local Similarity 7.7%; Score 82.5; DB 1; Length 2483;
Matches 39; Conservative 32; Mismatches 49; Indels 41; Gaps 7;

QY 45 TIDPMKVPDHADKFERHVGIVDFKGL-----AMRNIEAGLKQMKRG----- 88
DB 2235 TINPLCIEMAYADK-ESRGVLEEGEVTEIKFKREDIKMRIRDPAYKKLMELGSGPDS 2293
QY 89 DANWKGEEIVKRAHLIGVHDIYSMEYDLAYKLGDLPT-----THVISDIQDFVAL 142
DB 2294 DKDRKDLSEGRLLAR-----EDLLPIYHQVAVQAFPHDTGRLKRGVSIIDLEWKTAR 2348
QY 143 S-----LEISDEGNITMTSPEVROFANVYVNHIGLSIL 175
DB 2349 TELYWRRLRLLEDO-----VKQELIQASGELSHVITQSM 2384

RESULT 30
DIDH_CHLPN STANDARD; PRT; 461 AA.
AC Q92773; Q92086;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dihydrodipicolinate dehydrogenase (EC 1.8.1.4) (E3 component of 2-oxoglutarate dehydrogenase complex).
GN LPDA OR CPN0833 OR CP1037.
OS Chlamydia pneumoniae (Chlamydiophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=9920606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:365-369(1999).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RN Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RA Shirai M.;
RT "Genomic sequence comparison of two unrelated isolates of Chlamydia
RT pneumoniae from Japan and U.S.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMATIC COMPONENTS:
CC BRANCHED-CHAIN ALPHA-KETO DECARBOXYLASE (E1), LIPOAMIDE
CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3) (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Dihydrodipicolinate + NAD(+) = liponamide + NADH.
CC -!- COFACTOR: FAD (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE001663; AAD18970.1; -
DR EMBL: AE002261; AAF38812.1; -
DR EMBL: AP002548; BAA99041.1; -
DR EMBL: AB035943; BAA88651.1; -
DR HSSP: P14218; 1LPP.
DR PHCI-2DPAGE; Q92773; -.
DR TIGR: CP1037; -.
DR InterPro: IPR001327; FAD_Pyr_redox.
DR InterPro: IPR000815; Hg_reductase.
DR InterPro: IPR001100; Pyr_redox.
DR InterPro: IPR001033; Pyridine_redox_2.
DR InterPro: IPR004099; Pyr_redox_dlm.
DR Pfam: PF00070; Pyr_redox; 1.
DR Pfam: PF02852; Pyr_redox_dlm; 1.
DR PRINTS: PR00368; FADPNR.
DR PRINTS: PR00945; HGRDPTASE.
DR PRINTS: PR00411; PNDROTASEI.
DR PRINTS: PR00469; PNDROTASEII.
DR PRODOM: PD000139; FAD_Pyr_redox; 1.
DR PROSITE: PS00076; PYRIDINE_REDOX_1; 1.
KM Redox-active center: Glycolysis; Oxidoreductase; NAD; Flavoprotein;
KM FAD: Complete proteome.
FT NP_BIND 6 36 FAD (ADP PART) (PROBABLE).
FT DISULFD 42 47 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 294 304 FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT_SITE 436 436 BY SIMILARITY.

```



```

QY 72 AMRNIEARGLKMKROGDNVKGESIVKAHLIGVHD---IVSMEDYLA----- 120
DB 157 TVGKLIAMKDV-----GKEGVITVEDGTGLDELVAEVEGQFDRGLSPYFINK 207
QY 121 -----KLGDLHPPTHYISIDQDFVALSLSEISDEGNITMTSEFV 159
DB 208 PETGAVELESPILLADKKISINIREALPYLEAVAKAGPDLITAEVGEALATL----- 262
QY 160 ROFANVNHIGL-----SILDPIFGVLSVDVLAIFODYVRKEMTVLAPAFKRELEKN 213
DB 263 -----VNTMRGIVKVAANKAPFG---DRKKAMLDIATLTAGTVISEIGLELEKS 312

RESULT 28
CH60_SERFI STANDARD: PRT: 540 AA.
AC 066204:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOGA OR GROEL.
OS Serratia ficaria.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=61651;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 1241;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RT J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB008144; BAA25221.1; -
DR HSSP: P06139; 1GRU.
DR InterPro: IPR001844; Chaperin_Cpn60.
DR InterPro: IPR002423; Cpn60/TCF-1.
DR Pfam: PF00118; cpn60_TCF1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCF1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
FT NON_TER 540 540
SQ SEQUENCE 540 AA: 56627 MW: CEA489AD395EED93 CRC64;

Query Match 7.7%; Score 82.5; DB 1; Length 540;
Best Local Similarity 18.1%; Pred. No. 23;
Matches 43; Conservative 44; Mismatches 81; Indels 69; Gaps 9;

```

```

QY 121 -----KLGDLHPPTHYISIDQDFVALSLSEISDEGNITMTSEFV 159
DB 208 PETGAVELESPILLADKKISINIREALPYLEAVAKAGPDLITAEVGEALATL----- 262
QY 160 ROFANVNHIGL-----SILDPIFGVLSVDVLAIFODYVRKEMTVLAPAFKRELEKN 212
DB 263 -----VNTMRGIVKVAANKAPFG---DRKKAMLDIATLTAGTVISEIGLELEK 311

RESULT 29
COA2_HUMAN STANDARD: PRT: 2483 AA.
AC 000763; Q16852;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetyl-CoA carboxylase 2 (EC 6.4.1.2) (ACC-beta) [Includes: Biotin
DE carboxylase (EC 6.3.4.14)].
GN ACACB OR ACC2 OR ACGB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RC TISSUE=Liver;
RX MEDLINE=97256787; PubMed=9099716;
RA Abu-Elheiga L., Almaraz-Ortega D.B., Baldini A., Wakil S.J.;
RT "Human acetyl-CoA carboxylase 2. Molecular cloning, characterization,
RT chromosomal mapping, and evidence for two isoforms."
RT J. Biol. Chem. 272:10669-10677(1997).
RN [2]
RP SEQUENCE OF 1349-2134 FROM N.A.
RC TISSUE=Adipose tissue;
RX MEDLINE=96265061; PubMed=8670171;
RA Widmer J., Fassini K.S., Schlichter S.C., Wheeler K.S., Crute B.E.,
RA King N., Nulle-mcmenemy N., Noll W.W., Daniel S., Ha J., Kim K.-H.,
RA Witters L.A.;
RT "Identification of a second human acetyl-CoA carboxylase gene."
RT Biochem. J. 316:915-922(1996).
CC -1- FUNCTION: ACC-BETA MAY BE INVOLVED IN THE PROVISION OF MALONYL-COA
CC OR IN THE REGULATION OF FATTY ACID OXIDATION, RATHER THAN FATTY
CC ACID BIOSYNTHESIS. THIS PROTEIN CARRIES THREE FUNCTIONS: BIOTIN
CC CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE, AND
CC CARBOXYLTRANSFERASE.
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) - ADP + phosphate
CC + malonyl-CoA.
CC -1- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + Co(2)
CC = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.
CC -1- COFACTOR: BIOTIN.
CC -1- PATHWAY: Long-chain fatty acid biosynthesis; first (rate limiting)
CC step.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; MAY ASSOCIATE WITH MEMBRANES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE HEART, SKELETAL
CC MUSCLES AND LIVER.
CC -1- SIMILARITY: PARTIAL TO CARBOXYL PHOSPHATE SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09344; AAB58382.1; -
DR EMBL: U34591; AAC50571.1; -
DR HSSP: P24182; IDV1.
DR Genew, HGNC:85; ACACB.
DR InterPro: IPR001882; Biotin_attach.

```

KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW Complete proteome.
 SQ SEQUENCE 1060 AA; 121571 MW; 44803F0937A8D3E6 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 1060;
 Best Local Similarity 21.2%; Pred. No. 47;
 Matches 32; Conservative 29; Mismatches 38; Indels 52; Gaps 8;

QY 91 NVKGE-----EGIVKAHLIGVHDDIVSMEDLAYK-----LGDLHPPTHTVHSIDQDF-- 138
 DB 6 NKTETSELDSDVYKV-----DYLETAHRLGTYGTGCDVGNLHAAFRVRAKQKREVL 58
 QY 139 --VVALSEISDEG-----NT-----TMTSEFVROFANVNVHIGSLTSD 176
 DB 59 QPITSEIENFEWRGPIAFSFIADKTEGYNKLLRISTLHNVGROQFSDIONHLSGIALIT 118
 QY 177 P-IFGVLSDVLTAFIDTVRKEMTKVLAAPF 206
 DB 119 PETYGSLSL-----TELSVADAEAF 139

RESULT 26

CH60_ENTAM STANDARD; PRT; 539 AA.
 AC 066196;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
 GN GROU. OR MOPA OR GROEL.
 OS Enterobacter amylgenus.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Enterobacter.
 OX NCBI_TaxID=61646;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 1237;
 RA Harada H., Ishikawa H.;
 RT "Phylogenetical relationship based on groE genes among phenotypically related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia species."
 RL J. Gen. Appl. Microbiol. 43:355-361(1997).
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB008140; BAA25213.1; -
 DR HSSP: P06139; 1JON.
 DR InterPro: IPR001844; Chaprinin_Cpn60.
 DR InterPro: IPR002423; Cpn60/TCF-1.
 DR Pfam: PF00118; Cpn60_TCF1; 1.
 DR PRINTS: PR00298; CHAPERONIN60.
 DR PRINTS: PR00304; TCOMPLEXTCPL.
 DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
 KM Chaperone; ATP-binding.
 FT NON_TER 539 539
 SQ SEQUENCE 539 AA; 56657 MW; 09E701CB2B1DBA17 CRC64;

Query Match 7.7%; Score 82.5; DB 1; Length 539;
 Best Local Similarity 18.6%; Pred. No. 23;
 Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIEQSETIDPKKVPDHPADKFERHGVDFKQEL 71
 DB 107 VAAGNMFM-----DLKRGIDKAVIAAVEELKT---LSVPQSDSKAIAQGTISANSDE 156
 QY 72 AMRNIETARGLKOMKROGANVNGEGIVKAHLITGVHD---IYSMEDLAY----- 120
 DB 157 TVGKLIAEMOKV-----GKEGVITVEDGTLEDELVDVEQMQRDGLSPYFINK 207
 QY 121 -----KLGDLPHTHTVHSIDQFVVALSEISDEGNITMTSEFV 159
 DB 208 PETGAVELESPIILLADKKISIRREMLVLEAVAKGAPLLIADVGEALATL----- 262
 QY 160 ROFANVNVHIGSL-----SILDPTEGVLSDVLTAFIDTVRKEMTKVLAAPKRELEK 212
 DB 263 -----VVMTRMGIVAVAAVKAPGFG---DRKRAMQDLATLTLGCVISEELGMELEK 311

RESULT 27

CH60_KLEPL STANDARD; PRT; 540 AA.
 AC 066212;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
 GN GROU. OR MOPA OR GROEL.
 OS Klebsiella planticola.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Raoultella.
 OX NCBI_TaxID=575;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 7251;
 RA Harada H., Ishikawa H.;
 RT "Phylogenetical relationship based on groE genes among phenotypically related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia species."
 RL J. Gen. Appl. Microbiol. 43:355-361(1997).
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB008148; BAA25229.1; -
 DR HSSP: P06139; 1JON.
 DR InterPro: IPR001844; Chaprinin_Cpn60.
 DR InterPro: IPR002423; Cpn60/TCF-1.
 DR Pfam: PF00118; Cpn60_TCF1; 1.
 DR PRINTS: PR00298; CHAPERONIN60.
 DR PRINTS: PR00304; TCOMPLEXTCPL.
 DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
 KM Chaperone; ATP-binding.
 FT NON_TER 540 540
 SQ SEQUENCE 540 AA; 56655 MW; 9E18C1ABC2430A CRC64;

Query Match 7.7%; Score 82.5; DB 1; Length 540;
 Best Local Similarity 18.1%; Pred. No. 23;
 Matches 43; Conservative 45; Mismatches 81; Indels 69; Gaps 9;
 QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIEQSETIDPKKVPDHPADKFERHGVDFKQEL 71
 DB 107 VAAGNMFM-----DLKRGIDKAVIAAVEELKT---LSVPQSDSKAIAQGTISANSDE 156

DB 157 TGEALIAQAMEKV-----GKEGVITVEEGSGIQLDELVDVEGNOFDRGVISPYFINK 207
QY 121 -----KLGDLPHTTHVTSIDIQDFVALSLSEISDEGNITMTSFEV 159
DB 208 PENGSEIESPILLADKKISNIREMLPYLENAKAKGPLLIIAEDVEGALATL----- 262
QY 160 ROFANVNHIGL-----SLIDPIFGVLSVLTAIFODTVRKREMTKYLAFAKRELEK 212
DB 263 -----VVNTMRGIVKVAAVKAPGFG---DRKRAMLODIATLITAGVISEIGLELEK 311

RESULT 22

RS4E.META
ID RS4E.META STANDARD: PRT: 244 AA.
AC RS4E.META
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE 30S ribosomal protein S4e.
GN RPS4E OR M00468.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococci: Methanococcales;
OC Methanocaldococcaceae: Methanocaldococcus.
OX NCBI_TaxID=2150;
RN [1]

SEQUENCE FROM N.A.
RP STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geophagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.F., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RT Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE S4E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: 067497; AAB98457.1; -.
DR TIGR: M00468; -.
DR InterPro: IPR000302; K0W.motif.
DR InterPro: IPR000876; Ribosomal_S4E.
DR InterPro: IPR002942; S4.
DR Pfam: PF00467; K0W; 1.
DR Pfam: PF00900; Ribosomal_S4e; 1.
DR Pfam: PF01479; S4; 1.
DR ProDom: PD002667; Ribosomal_S4E; 1.
DR SMART: SM00363; S4; 1.
DR PROSITE: PS00528; RIBOSOMAL_S4E; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 244 AA: 27658 MW: 6759485019B86AC7 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 244;
Best Local Similarity 21.5%; Pred. No. 8;
Matches 42; Conservative 30; Mismatches 87; Indels 36; Gaps 6;

QY 6 LIAAVAFVVSADPIHYDKITEINKAIDAIAEQSEITIDPMKYPDHAKFERHVGIV 65
DB 45 LLIVVDILKYNDAEAKKIIMKGVLYDGRYRKREKLPVGLMDVSLPDANENRVLF 104
QY 66 DEKGEIAMRNIEARGLKQKKROGDANVKGEGIVAKHLLIGVHD-----DIVSMEX 116

DB 105 DRKRIRIKLPENPDVKICKIKNKVIKG-----GHQLNLHNGRNIVIKVSDPTKAE 158
QY 117 DLAVKIGDLHPHTTHVTSIDIQDFVALSLSEISDEGNITMTSFEVROFANV--NHIGLSI 174
DB 159 DV-YKTGD-----TLLISIPDEIRKAHLPFEGKGLAYITGKRHVDPAK 201
QY 175 LDPI--FGVLSVLTA 187
DB 202 IVERERRGIYPIVIT 216

RESULT 23

APEA_CLOAB
ID APEA_CLOAB STANDARD: PRT: 465 AA.
AC APEA_CLOAB
DT 15-JUN-2002 (rel. 41, Created)
DT 15-JUN-2002 (rel. 41, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Probable M18-family aminopeptidase 1 (EC 3.4.11.-).
GN APEA OR CAC1091.
OS Clostridium acetobutylicum.
OC Bacteria: Firmicutes: Clostridia; Clostridiales: Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]

SEQUENCE FROM N.A.
RP STRAIN-ATCC 824 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
RA Tatusov R.L., Sabath F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RT Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RT J. Bacteriol. 183:4823-4838(2001).
CC -1- COFACTOR: Zinc (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M18.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AE007625; AAK79065.1; -.
DR InterPro: IPR001948; Aminolipase.
DR Pfam: PF02127; Peptidase_M18; 1.
DR PRINTS: PR00932; AMINOPEPTASE.
KW Hypothetical protein; Hydrolase; Aminopeptidase; Metalloprotease;
KW Zinc; Complete proteome.
FT METAL 105 105 ZINC (POTENTIAL).
FT METAL 180 180 ZINC (POTENTIAL).
FT METAL 441 441 ZINC (POTENTIAL).
SQ SEQUENCE 465 AA: 52011 MW: CBAC6D0A54C0A439 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 465;
Best Local Similarity 28.4%; Pred. No. 17;
Matches 42; Conservative 27; Mismatches 49; Indels 30; Gaps 10;

QY 74 RNE---ARGLKQMKRQGO---ANVK-----GEEGIVAKHLLITGVHDI--VSMEX 116
DB 58 RNIEDILAKG--ETLKEGKQVYANNNGKGLIMFLICKREPLYTGKILGHIDSPRLDKQ 115
QY 117 DLAVKIGDLHP--THVTSIDIOFV--VALSLSEISDEGNITMTSFEVROFANVNHIGLSI 174
DB 116 NPLVETDPLAMELTHYGGIKKYOWTLPLAI--HGVI-----YAKDGTIVNVCGEED 167
QY 175 LDPIFGVLSVLTAIQDTPVREMTYVL 202
DB 168 NDPVFGV--SDIIVHLASLEQLERKASKVI 194

Query Match	7.9%	Score 84:	DB 1:	Length 358:
Best Local Similarity	23.4%	Pred. No. 11:		
Matches	33:	Conservative	26:	Mismatches 50: Indels 32: Gaps 5:
QY	27	EEINAAIDDAIAAPOSSETIDPMKVPDHA	DR	EMBL, AL031541; CAA20806.1; ALT_INIT.
DB	4	EEIEMRDEBALAFPAASDLDLQAKVAHNGGASPLALANREITALLPQAAAEAGKRVG	DR	HSSP: P27001. 1PVS.
QY	77	EARGL-----KMKRGDANKVGEGYIKAKILLIGVHDDIYSMEYDLAYKKGDLHP	DR	InterPro: IPR002106; AATRNA_ligase1.
DB	64	MARGAVNKAALAAQGELEARDARLVLEAV-----DVTLPND-RVPAGARHP	DR	InterPro: IPR004529; PHEs.
QY	128	TTVAISDIQDFVVALSLEISD 148	DR	InterPro: IPR004188; PHE-trNA_synt_N.
DB	111	LTTLSERIEDIEFVAMGYEVAE 131	DR	InterPro: IPR002319; trNA_synt_2d.
			DR	Pfam: PF01409; trNA_synt_2d. 1.
			DR	Pfam: PF02912; PHE-trNA_synt_N. 1.
			DR	TIGRFAMS: TIGR00468; PHEs. 1.
			DR	PROSITE: PS50862; AA-trNA_LIGASE_1.
			DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
			DR	SEQUENCE 358 AA; 39119 MW; 49CF66E0000E469C CRC64;
			DR	InterPro: IPR002106; AATRNA_ligase1.
			DR	InterPro: IPR004529; PHEs.
			DR	InterPro: IPR004188; PHE-trNA_synt_N.
			DR	InterPro: IPR002319; trNA_synt_2d.
			DR	Pfam: PF01409; trNA_synt_2d. 1.
			DR	Pfam: PF02912; PHE-trNA_synt_N. 1.
			DR	TIGRFAMS: TIGR00468; PHEs. 1.
			DR	PROSITE: PS50862; AA-trNA_LIGASE_1.
			DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
			DR	SEQUENCE 358 AA; 39119 MW; 49CF66E0000E469C CRC64;
			DR	InterPro: IPR002106; AATRNA_ligase1.
			DR	InterPro: IPR004529; PHEs.
			DR	InterPro: IPR004188; PHE-trNA_synt_N.
			DR	InterPro: IPR002319; trNA_synt_2d.
			DR	Pfam: PF01409; trNA_synt_2d. 1.
			DR	Pfam: PF02912; PHE-trNA_synt_N. 1.
			DR	TIGRFAMS: TIGR00468; PHEs. 1.
			DR	PROSITE: PS50862; AA-trNA_LIGASE_1.
			DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
			DR	SEQUENCE 358 AA; 39119 MW; 49CF66E0000E469C CRC64;
			DR	InterPro: IPR002106; AATRNA_ligase1.
			DR	InterPro: IPR004529; PHEs.
			DR	InterPro: IPR004188; PHE-trNA_synt_N.
			DR	InterPro: IPR002319; trNA_synt_2d.
			DR	Pfam: PF01409; trNA_synt_2d. 1.
			DR	Pfam: PF02912; PHE-trNA_synt_N. 1.
			DR	TIGRFAMS: TIGR00468; PHEs. 1.
			DR	PROSITE: PS50862; AA-trNA_LIGASE_1.
			DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
			DR	SEQUENCE 358 AA; 39119 MW; 49CF66E0000E469C CRC64;
			DR	InterPro: IPR002106; AATRNA_ligase1.
			DR	InterPro: IPR004529; PHEs.
			DR	InterPro: IPR004188; PHE-trNA_synt_N.
			DR	InterPro: IPR002319; trNA_synt_2d.
			DR	Pfam: PF01409; trNA_synt_2d. 1.
			DR	Pfam: PF02912; PHE-trNA_synt_N. 1.
			DR	TIGRFAMS: TIGR00468; PHEs. 1.
			DR	PROSITE: PS50862; AA-trNA_LIGASE_1.
			DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
			DR	SEQUENCE 358 AA; 39119 MW; 49CF66E0000E469C CRC64;
			DR	InterPro: IPR002106; AATRNA_ligase1.
			DR	InterPro: IPR004529; PHEs.
			DR	InterPro: IPR004188; PHE-trNA_synt_N.
			DR	InterPro: IPR002319; trNA_synt_2d.
			DR	Pfam: PF01409; trNA_synt_2d. 1.
			DR	Pfam: PF02912; PHE-trNA_synt_N. 1.
			DR	TIGRFAMS: TIGR00468; PHEs. 1.
			DR	PROSITE: PS50862; AA-trNA_LIGASE_1.
			DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
			DR	SEQUENCE 358 AA; 39119 MW; 49CF66E0000E469C CRC64;
			DR	InterPro: IPR002106; AATRNA_ligase1.
			DR	InterPro: IPR004529; PHEs.
			DR	InterPro: IPR004188; PHE-trNA_synt_N.
			DR	InterPro: IPR002319; trNA_synt_2d.
			DR	Pfam: PF01409; trNA_synt_2d. 1.
			DR	Pfam: PF02912; PHE-trNA_synt_N. 1.
			DR	TIGRFAMS: TIGR00468; PHEs. 1.
			DR	PROSITE: PS50862; AA-trNA_LIGASE_1.
			DR	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
			DR	SEQUENCE 358 AA; 39119 MW; 49CF66E0000E469C CRC64;
			DR	InterPro: IPR002106; AATRNA_ligase1.
			DR	InterPro: IPR004529; PHEs.
			DR	InterPro: IPR004188; PHE-trNA_synt_N.
			DR	InterPro: IPR002319; trNA_synt_2d.
			DR	

```

ID      XYLG_ECOLI      STANDARD:      PRT:      513 AA.
DT      P37388;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      D-xylose transport ATP-binding protein xylG.
CN      XYLG OR B3567.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655.
RX      MEDLINE=94316500; Pubmed=8041620;
RA      Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT      "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT      region from 76.0 to 81.5 minutes."
RL      Nucleic Acids Res. 22:2576-2586(1994).
CC      -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY D-XYLOSE MEMBRANE
CC      TRANSPORT SYSTEM.
CC      -1- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).
CC      -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; U000039; AB18544.1; -.
DR      EMBL; AE000434; AAC76591.1; -.
DR      HSSP; Q58663; 1G6H.
DR      COSSEP; EG12275; xylG.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transportr.
DR      Pfam; PF00005; ABC_tran; 2.
DR      ProDom; PD000006; ABC_transportr; 2.
DR      SMART; SMO0382; AAA; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW      Transport; Sugar transport; Inner membrane; ATP-binding;
KW      Complete proteome.
KW      NP_BIND
KW      FT
SQ      SEQUENCE 513 AA; 56470 MM; FAD2C65D178D50DF CRC64;
      37 44
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
      61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
      91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 
```

RC STRAIN-Berkeley;
 RA MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abill J.F., Abayanti A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Baller M.R., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies S.M.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Durkin K.J., Evans J.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibeigbam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Modyarty C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden Klamos I., Simpson M., Skupsky M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP SEQUENCE OF 1-27 FROM N.A.
 RX MEDLINE-87089662; PubMed-3025586;
 RA Seales L.L., Greenleaf A.L., Kemp W.E., Voelker R.A.;
 RT "Sites of P element insertion and structures of P element deletions in
 RT the 5' region of *Drosophila melanogaster* Rpr1215.";
 RL Mol. Cell. Biol. 6:3312-3319(1986).
 RN [4]
 RP SEQUENCE OF 1-472 FROM N.A.
 RX MEDLINE-85282618; PubMed-2992806;
 RA Biggs J., Seales L.L., Greenleaf A.L.;
 RT "Structure of the eukaryotic transcription apparatus: features of the
 RT gene for the largest subunit of *Drosophila* RNA polymerase II.";
 RL Cell 42:611-621(1985).
 RN [5]
 RP SEQUENCE OF 1441-1887 FROM N.A.
 RX MEDLINE-88094402; PubMed-3122024;
 RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.;
 RT "The C-terminal domain of the largest subunit of RNA polymerase II of
 RT Saccharomyces cerevisiae, *Drosophila melanogaster*, and mammals: a
 RT conserved structure with an essential function.";
 RL Mol. Cell. Biol. 8:321-329(1988).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO
 CC TWELVE DIFFERENT POLYPEPTIDES. THE 215 KDA POLYPEPTIDE IS THE
 CC LARGEST COMPONENT OF RNA POLYMERASE II.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
 CC THE PHOSPHORYLATION ACTIVATES POL2.

CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND 5.8S GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M27431; AAA28868.1; -
 CC DR EMBL: AE003486; AAF48057.1; -
 CC DR EMBL: M14203; AAA28864.1; -
 CC DR EMBL: M11798; AAA28863.1; -
 CC DR EMBL: M19537; AAA28827.1; -
 CC DR PIR: S04457; RNF2L.
 CC DR FLYBASE: FBgn0003277; Rpl1215.
 CC DR InterPro: IPR000684; RNA_polII_repeat.
 CC DR InterPro: IPR000722; RNA_pol_A.
 CC DR InterPro: IPR002879; RNA_pol_A2.
 CC DR Pfam: PF00623; RNA_pol_A; 1.
 CC DR Pfam: PF01854; RNA_pol_A2; 1.
 CC DR PROSITE: PS00115; RNA_pol_II_REPEAT; 11.
 CC KW Transferring: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
 CC DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
 CC FT ZN-FING 67 83
 CC FT DOMAIN 1579 1881
 CC FT CONFLICT 319 324 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
 CC FT CONFLICT 450 450 F -> G (IN REF. 4).
 CC FT CONFLICT 455 458 TLHK -> RCTT (IN REF. 4).
 CC FT CONFLICT 463 472 GHRYVLPSS -> VTGEVASST (IN REF. 4).
 CC FT CONFLICT 741 741 R -> H (IN REF. 1).
 CC FT CONFLICT 1485 1524 SMLGAAFTIGGSPSPMPPTWPNACNTPTTFSPGHH
 CC FT CONFLICT 1506 1508 MTP -> EIDSA (IN REF. 1).
 CC FT CONFLICT 1887 1887 D -> DVKRGGRG (IN REF. 1).
 CC FT CONFLICT 1887 1887 D -> DVKRGGRG (IN REF. 1).
 CC SQ SEQUENCE 1887 AA; 209167 MW; 4EC68C7708A167A3 CRC64;
 CC -----
 CC Query Match 7.9%; Score 84.5; DB 1; Length 1887;
 CC Best Local Similarity 22.7%; Pred. No. 71;
 CC Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12;
 CC -----
 CC QY 17 ADPIHDKITEENKAIDAIAIEOSERTIDPKAVYDHA--DKFEHNYGV-----DFRG 69
 CC DB 686 ADPQYNEIQDAIKKAKDDVINIQKHNMLEPPTGNTLRQTFENKKNVRLINDARDKGT 745
 CC QY 70 ELAMRNI-EARGLKKMRQ-----GDANYGEE--GIYKAHLIGVHD 109
 CC DB 746 GSAKKSILTEYNNLKAANYSSKSGSNINISYIVACVQOQVNEGRIYGFPRKRLPLPFIND 805
 CC QY 110 D-----IVSMEDLAYKIGDLHPTT---HWISDIQDFVALSLEISDEGNI-----T 153
 CC DB 806 DYPESRGFEVENS-----LAGLTPESEYFHAMGG-REGILIDPAVTAERGVIQRRILKA 859
 CC QY 154 MTSFEYRQFANVYNNHIGLSIL-----DPIFGVLS----- 184
 CC DB 860 MESVMVNYDQTVNSVSGQLQIRYGEDGLGELVEFQNMPTVYLSNKSFEKRFKFDMSNE 919
 CC QY 185 -VLTALFQDFTREKMT 199
 CC DB 920 RLKKVYFTDVIKEMT 935
 CC -----
 CC RESULT 18
 CC SYFA_STRCO STANDARD; PRT; 358 AA.
 CC AC 088055;
 CC DT 16-OCT-2001 (rel. 40, Created)
 CC 16-OCT-2001 (rel. 40, Last sequence update)

RC STRAIN-JCM 6096;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB008149; BAA25231.1; -.
DR HSP; P06139; 1JON.
DR InterPro: IPR001844; Chaprinln_Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
DR Chaperone; ATP-binding.
KW NON_TER
FT 540
SQ SEQUENCE 540 AA; 56361 MW; 6DB51337F5C30D8C CRC64;

Query Match 7.9%; Score 84.5; DB 1; Length 540;
Best Local Similarity 18.6%; Pred. No. 16;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAIEQSETIDPMKVPPOHADKFEHVGIVDFKGL 71
DB 107 VAAGNMPM-----DLKRGIDKAVVAVAEELKT--LSVPCDSKAIAQVGTISANSDE 156
QY 72 AMRNIARLQKKRQGDANVGEIGIVAKHLIGVHD--IVSMEDYLA-----120
DB 157 TVGKLIAMDKV-----GKEGVITVEDGTGLEDELVDVEGQFGRGLSPFYINK 207
QY 121 -----KLGDLPPTTHVISDIQDFVVALSLTISDEGNITMTSFEV 159
DB 208 PETGAVELESPFLLADKKVSNIREMLPVLEAVAKGKPLVIAEDVEGEALATL-----262
QY 160 ROFANVNHIGGL-----SLDPIFGVLSVLAIFQDITRKEMTKYLAFAFKRELEK 212
DB 263 -----VVNTMRGIVKVAVAKAPGFG--DRKKAMQDIAITLGTGIVISEIGMELEK 311
RESULT 16
CH60_KLEOX
ID CH60_KLEOX STANDARD; PRT; 540 AA.
AC 066210;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR MOXA OR GROEL.
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-JCM 1665;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia

RT species";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB008147; BAA25227.1; -.
DR HSP; P06139; 1JON.
DR InterPro: IPR001844; Chaprinln_Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
DR Chaperone; ATP-binding.
KW NON_TER
FT 540
SQ SEQUENCE 540 AA; 56445 MW; BFE441BC6B05E06B CRC64;

Query Match 7.9%; Score 84.5; DB 1; Length 540;
Best Local Similarity 18.6%; Pred. No. 16;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAIEQSETIDPMKVPPOHADKFEHVGIVDFKGL 71
DB 107 VAAGNMPM-----DLKRGIDKAVVAVAEELKT--LSVPCDSKAIAQVGTISANSDE 156
QY 72 AMRNIARLQKKRQGDANVGEIGIVAKHLIGVHD--IVSMEDYLA-----120
DB 157 TVGKLIAMDKV-----GKEGVITVEDGTGLEDELVDVEGQFGRGLSPFYINK 207
QY 121 -----KLGDLPPTTHVISDIQDFVVALSLTISDEGNITMTSFEV 159
DB 208 PETGAVELESPFLLADKKISNIREMLPVLEAVAKGKPLVIAEDVEGEALATL-----262
QY 160 ROFANVNHIGGL-----SLDPIFGVLSVLAIFQDITRKEMTKYLAFAFKRELEK 212
DB 263 -----VVNTMRGIVKVAVAKAPGFG--DRKKAMQDIAITLGTGIVISEIGMELEK 311
RESULT 17
RPB1_DROME
ID RPB1_DROME STANDARD; PRT; 1887 AA.
AC P04052; Q9VYX6;
DT 01-NOV-1986 (Rel. 03, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).
GN RPII215 OR CG1534.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89218930; PubMed=2496296;
RA Jucker R.S., Weeks J.R., Zehring W.A., Greenleaf A.L.;
RT "Analysis of the gene encoding the largest subunit of RNA polymerase
RT II in Drosophila."
RL Mol. Genet. 215:266-275(1989).
RN [2]
RP SEQUENCE FROM N.A.


```

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/TW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001304; AAC67907.1; -.
DR HSSP: Q9KMW6; 1HQM.
DR PHCI-2DPAGE: 084316; -.
DR InterPro: IPR000722; RNA_pol_A.
DR InterPro: IPR002879; RNA_pol_A2.
DR Pfam: PF00623; RNA_pol_A; 1.
DR Pfam: PF01854; RNA_pol_A2; 2.
DR Transferrase: DNA-directed RNA polymerase; Transcription;
KW Complete proteome.
SQ SEQUENCE 1396 AA; 154904 MW; B24BF841D284065 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 1396;
Best Local Similarity 24.0%; Pred. No. 41;
Matches 50; Conservative 33; Mismatches 66; Indels 59; Gaps 12;

OY 43 SEFTIDPMKYVDHADRERHVGIVDFEGELAMRYI-----EARGLKO-----MKRQ 87
DB 1145 AELVERARKEPDAD-TAKIDGVVDFKGIQKNRILVROEITGMEBEHLISLTKHLIVOR 1203
OY 88 GDANVAGE---EGIVKAHLI---GVHD--DIYSMEYDLAYKLGDLHPTTHVISIDIOFV 139
DB 1204 GDSVYKGGQULTGVLVPHHLEICGRELQKYLVNEGVGYRQGV-----DINDKH 1255
OY 140 VAL-----SLEISDEGNITM---TSFEVROFANY---VNHIG-----LSILD 177
DB 1256 VEIIVRQMLQKRYITDPGCTTLTFLGEDVDKKEEYENRRTEEDGKPAQAVPVLGITKA 1315
OY 178 IFGVLSDVLTAFODTVRKEMTKYLAPA 205
DB 1316 SLGTESFISAAASFQDT-----TRVLTA 1338

RESULT 14
CH60_SERRU STANDARD; PRT; 539 AA.
AC 066202;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
OS GROL OR MOBA OR GROEL.
OC Serratia rubidaea.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```

OC Serratia.
OX NCBI_TaxID=61652;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 1240;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB008143; BAA25219.1; -.
DR HSSP: P06139; 1GRL.
DR InterPro: IPR001844; Chaperon_Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
FT NON_TER 539 539.
SQ SEQUENCE 539 AA; 56640 MW; CE447B6571504F0 CRC64;

Query Match 7.9%; Score 84.5; DB 1; Length 539;
Best Local Similarity 18.6%; Pred. No. 16;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

OY 13 VAVSADPIHYDKITEEINKAIDA-IAAIEQSEITDPMKVPDHADKFERHVGIVDFEGEL 71
DB 107 VAAGMNP-----DLKRGIDKAVIAAIVE--LTKLSVPCSDSKAIAQVGTISANSDE 156
OY 72 AARNITARBLKQMKRGDANVAGEEGIVKAHLIGVHD--IYSMEYDLA----- 120
DB 157 TVGKLIAEMERK-----GKEGVITVEEGTGLQDELDDVVEGQFDRGYLSPYFINK 207
OY 121 -----KLGDLHPTTHVISIDIOFVVALSLEISDEGNITWTSFEV 159
DB 208 PEGSVLESPFILLADKKVSNRELPPVLEAVAKKAPKLLIYAEBVEGALATL----- 262
OY 160 ROFANYVNHIGL-----SILDPIFVLSDVLTAFODTVRKEMTKYLAPAFKRELEK 212
DB 263 -----VVNTMRGIVKVAAVKAPQFG---DRRKAMLDIATLTAGTVISEIGMELEK 311

RESULT 15
CH60_KLEON STANDARD; PRT; 540 AA.
AC 066214;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
OS Klebsiella ornithinolytica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Raoultella.
OX NCBI_TaxID=54291;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodde A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klek H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -! SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES. P97/CDC48
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67557; AAB99153.1; -.
DR TIGR: MJ1156; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003959; AAA_ATPase_cent.
DR InterPro: IPR003960; AAA_sub.
DR InterPro: IPR003338; ATPaseVAT_N.
DR InterPro: IPR004201; Cdc48_2.
DR Pfam: PF00004; AAA; 2.
DR Pfam: PF03359; cdc48_N; 1.
DR Pfam: PF02933; cdc48_2; 1.
DR SMART: SM00382; AAA; 2.
DR TIGRFAMS: TIGR01243; CDC48; 1.
DR PROSITE: PS00674; AAA; 2.
KW Hypothetical protein; ATP-binding; Repeat; Multigene family;
KW Complete proteome.
FT NP_BIND 220 227 ATP (POTENTIAL).
FT NP_BIND 493 500 ATP (POTENTIAL).
SQ SEQUENCE 903 AA; 100403 MW; 99F2ABEE2D544DC2 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 903;
Best Local Similarity 25.0%; Pred. No. 24;
Matches 53; Conservative 30; Mismatches 90; Indels 39; Gaps 11;

QY 23 DKITEINKAIDD-----AIAIEQSETIDPMKV-PDHAKEFERHGIYDFKGLAM 73
DB 571 DKVAVQLLELDGMEPRQVVIATNRRDIIIPALLRGRDRVIL-VPVDEKARLDI 629
QY 74 RNIERAGLKQMKRGDANV---KGEGLVKAHL-----LIGVHDIYMEYDLAYK 121
DB 630 FKIHRR---SMNLADVNLLEELAKKTEGYTGADIALCREAMLAVERSI-GKPMIDIEVK 685
QY 122 LGDHPHTHVISDIDFVAVSLSEISDEGNIT--MTSPFVRQFANVNIIGSLIDPIF 179
DB 666 LREL--INVLOSISGTRAAAVELNSVIKATKERSAAGESELSKNAIG-----KII 736
QY 180 GVLSDVLTAFQDTPVKKEMTKVLAPAFKRELE 211
DB 737 SVLSPAKKEKI--EAVEKEIDKFLVINKELK 766

RESULT 12
RPOC_CHLMU STANDARD: PRT: 1396 AA.
AC 09PK79;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR TC0588.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;

```

```

RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN-MOPN / N199;
RC MEDLINE=20150255; PubMed=10684935;
RX Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolony J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -! FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -! CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -! SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -! SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002327; AAF39420.1; -.
DR HSSP: Q9KW06; 1HQW.
DR TIGR: TC0588; -.
DR InterPro: IPR00722; RNA_pol_A.
DR InterPro: IPR002879; RNA_pol_A2.
DR Pfam: PF00623; RNA_pol_A; 1.
DR Pfam: PF01854; RNA_pol_A2; 2.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Complete proteome.
SQ SEQUENCE 1396 AA; 154897 MW; CC090B8A188B6977 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 1396;
Best Local Similarity 23.6%; Pred. No. 41;
Matches 49; Conservative 34; Mismatches 66; Indels 59; Gaps 12;

QY 43 SETIDPMKPRDADKEFRHVGIVDFKGLAMNI-----EAGIKO-----MKRQ 87
DB 1145 AEIVEARKEPDAD--IAKIDGVVDFGIGQKNRILVVRDEVGMEEHLISLTKHLIVQR 1203
QY 88 GNAVYGE--EGIVKAHLI---GVHD--DIVSMFYDLAYKIGDLPTHTVYSIDQFV 139
DB 1204 GDSVIKGOQLTGGLVPHPHILELCGYRELQKVLNVEQVYRLQV-----DINMKH 1255
QY 140 VAL-----SLEISDEGNITM--TSFEVRQFANV--VNHIG-----LSIDP 177
DB 1256 IEIIVRQMLQKVRITDPDGTTLFGEDVDKKEFYENRRTERRDGKPAQAVPVLGITKA 1315
QY 178 IFGVLSDVLTAFQDTPVKKEMTKVLAPAFKRELE 205
DB 1316 SLGTESFISASFSQDT-----TRVLTA 1338

RESULT 13
RPOC_CHLTR STANDARD: PRT: 1396 AA.
AC 084316;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR CT314.
OS Chlamydia trachomatis.

```

```

Db 453 AEKQANI---ALAEIVNNKRLP-----DAKVAGNAVLEITKNDONTNLEKLM 503
Oy 105 IGVHDIYSME-----YDLAYKIGDLPHPTTHVSDIODEFVALSLTSDRG 150
Db 504 EATVALIINSELEKROKQMLEKAVDGLSKDASRAVITDGIKIDVYIKINLSDEK 563
Oy 151 NITWT---SEFROPANVNHIGLSILDPIFGVLSVDLTAIFQDTRKMTKVLAPAK 207
Db 564 TMLIANGVKDVNSELNAEKOKSLGSLVK--GVEKQVLSPPAQOOLIQHLDKITAQIK 621
Oy 208 RELK 212
Db 622 KDTIK 626

RESULT 10
FKB4_MOUSE STANDARD: PRT: 457 AA.
AC P30416:
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE FK506-binding protein 4 (possible peptidyl-prolyl cis-trans isomerase
DE FKB4) (EC 5.2.1.8) (PPIase) (p59 protein) (HSE binding
DE immunophilin) (HBI) (FKBP52 protein) (52 kDa FK506 binding protein)
DE (FKBP59).
GN FKB4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94040772; PubMed=7693550;
RA Schmitt J., Stuenkelberg H.G.;
RT Cloning and expression of a mouse cDNA encoding p59, an immunophilin
RT that associates with the glucocorticoid receptor."
RL Gene 132:267-271(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung tumor;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 5-457 FROM N.A., PARTIAL SEQUENCE, FUNCTION, AND
RP SUBCELLULAR LOCATION.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=93342084; PubMed=8341706;
RA Alnemir E.S., Fernandes-Alnemir T., Nelki D.S., Dudley K.,
RA Dubois G.C., Litwack G.;
RT "Overexpression, characterization, and purification of a recombinant
RT mouse immunophilin FKBP-52 and identification of an associated
RT phosphoprotein."
RL Proc. Natl. Acad. Sci. U.S.A. 90:6839-6843(1993).
CC -I- FUNCTION: Component of unactivated mammalian steroid receptor
CC complexes that sediment at 8-10 S. May have a rotamase activity.
CC May play a role in the intracellular trafficking of hetero-
CC oligomeric forms of steroid hormone receptors.
CC -I- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -I- SUBUNIT: Associates with HSP70 and HSP70 in unactivated steroid
CC hormone receptor complexes. Also interacts with peroxisomal
CC phytanoyl-coa alpha-hydroxylase (PHYH) (By similarity).
CC -I- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -I- PTM: Phosphorylation by CK2 results in loss of HSP90 binding
CC activity (By similarity).
CC -I- SIMILARITY: BELONGS TO THE FKB4-TYPE PPIASE FAMILY.
CC -I- SIMILARITY: CONTAINS 2 FKB4-TYPE PPIASE DOMAINS.
CC -I- SIMILARITY: CONTAINS 3 TPR REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X70887; CAA50231.1; -.
CC EMBL: BC003447; AAH03447.1; -.
CC EMBL: X17069; CAC39452.1; -.
CC EMBL: X17068; CAA34914.1; ALT_SEQ.
CC HSSP: P27124; 1ROT.
CC MGD: MGI:95543; Fkbp4.
CC InterPro: IPR001179; FKBP_PPIase.
CC InterPro: IPR001440; TPR.
CC Pfam: PF00254; FKBP; 2.
CC Pfam: PF00515; TPR; 3.
CC SMART: SM00028; TPR; 3.
CC PROSITE: PS00453; FKBP_PPIASE_1; 1.
CC PROSITE: PS00454; FKBP_PPIASE_2; 2.
CC PROSITE: PS50059; FKBP_PPIASE_3; 2.
CC K1W Isomerase; Rotamase; Repeat; TPR repeat; Nuclear protein;
CC Phosphorylation.
CC INT_MET 0 0
CC DOMAIN 49 137 BY SIMILARITY.
CC DOMAIN 166 252 PPIASE, FKBP-TYPE 1.
CC REPEAT 269 302 PPIASE, FKBP-TYPE 2.
CC REPEAT 318 351 TPR 1.
CC REPEAT 352 385 TPR 2.
CC MOD_RES 142 142 TPR 3.
CC MOD_RES 142 142 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
CC CONFLICT 5 6 MK -> HE (IN REF. 3).
CC CONFLICT 155 162 TRGEQYAR -> LQYKMOG (IN REF. 3).
CC CONFLICT 202 239 GLEAIVORMERGESIVTLKPSVAGSGVGERFOIPIPH ->
CC AMRPFSAWRRESIPSTSNLAMLAVLWGRGSRSHRT
CC (IN REF. 3).
CC CONFLICT 264 271 EKLQSNL -> RSMRATY (IN REF. 3).
CC CONFLICT 314 314 H -> R (IN REF. 2).
CC SEQUENCE 457 AA: 51441 MW: 000875435410100 CRC64;

Query Match 8.1%; Score 86.5; DB 1; Length 457;
Best Local Similarity 23.0%; Pred. No. 8.9;
Matches 28; Conservative 27; Mismatches 50; Indels 17; Gaps 3;

Oy 18 DPHIDKITEELINKAIDDAIAIEOSEI-----DPMKVPDHD-KFERHY 62
Db 74 DKFSDLGKEVIAKMDAVATMKVGEVCHITCKREYAGAGSPKIPNATLVE--V 131
Oy 63 GIVDFKGLAMNIEARGIKOKKROGDANVKGEGIVKAKHLIGVHDIYSMEYDLAYKL 122
Db 132 ELPERFGEDELTEEDDGIIRITRKEGYARPNDCAMVEALLEGYHKRDLFQRELCEV 191
Oy 123 GD 124
Db 192 GE 193

RESULT 11
YB56_METJA STANDARD: PRT: 903 AA.
AC YB56_METJA
ID YB5656:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division cycle protein 48 homolog MLL156.
GN MLL156.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

```

```

OY 3 KFLIIAAVAVASADPIHDKITEINKAIDA---IAAIEOSE---TIDPMKVPDHA 55
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 KFAELTRINLEAKQVEQMLRDELADSVSKTYSDDADROHLELEKSEMELKVEVSLKEIS 238
OY 56 DKEFHNVGVDFKGLAMRNIEARGLKOMKROGDANVKGEEIVAHLLIGVHDIYSKE 115
    | : | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 239 DIARKQVEILNMQOOSRKEVESLRTOLDYQAOQSDERA-----LIAKLHQHVSIQ 290
OY 116 YDLAKYKGLDHPHTHVISDIODFVVALSLEISDE 149
DB 291 ASEAAALGKVESVASKLOKVEAHITRLQKLDKEP 324

RESULT 8
CH60_LEGPN STANDARD: PRT: 547 AA.
AC P26878.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (58 kDa common
DE antigen) (Heat shock protein B).
GN GROEL OR MOXA OR GROEL OR HTPB.
OS Legionella pneumophila.
OC Bacteria: Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90354095; PubMed=2117582;
RA Sampson J.S., O'Connor S.P., Holloway B.P., Plikaytis B.B.,
RA Carlone G.M., Mayer L.W.;
RT "Nucleotide sequence of htpB, the Legionella pneumophila gene
RT encoding the 58-kilodalton (kDa) common antigen, formerly designated
RT the 60-kDa common antigen."
RL Infect. Immun. 58:3154-3157(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90382960; PubMed=2205580;
RA Hoffman P.S., Houston L., Butler C.A.;
RT "Legionella pneumophila htpAB heat shock operon: nucleotide sequence
RT and expression of the 60-kilodalton antigen in L.
RT pneumophila-infected HeLa cells."
RL Infect. Immun. 58:3380-3387(1990).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE RENOVING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- FUNCTION: MAY PLAY A PROTECTIVE ROLE AGAINST THE DEFENSE
CC MECHANISMS GENERATED BY THE INFECTED MACROPHAGES.
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M91673; AAA25298.1; -.
DR EMBL: M31918; AAA25299.1; -.
DR PIR: A41472; A41472.
DR PIR: A41468; A41468.
DR HSSP: P06139; 1GRL.
DR InterPro: IPR001844; Chaperin_Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR Pfam: PF00118; cpn60_TCP1.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONIN5_CPN60; 1.

```

```

KW Chapterone: ATP-binding; Heat shock.
FT CONFLICT 187 187 S->Y (IN REF. 2).
FT CONFLICT 194 198 LIAVH->PGRGI (IN REF. 2).
FT CONFLICT 345 345 T->A (IN REF. 2).
SQ SEQUENCE 547 AA; 57946 MW; FA088F0F810EBEFC CnC64;

Query Match 8.2%; Score 88; DB 1; Length 547;
Best Local Similarity 21.3%; Pred. No. 8.4;
Matches 49; Conservative 50; Mismatches 97; Indels 34; Gaps 12;

OY 1 MKKFLIIAAVAVASADPIHDKITEINKAIDAIAIEOSEITIDPMKVPDHADEKER 60
DB 94 LARSILVEGHKRAVAGAMPN-----DKRGIDAVAVLAV--TKRLQMSRKCKDSKAIA 144
OY 61 HVGIVDFKGLAMRNIEARGLKOMKROG-----DAN-VKGEIGYKAHLLIGVHDD-IVS 113
DB 145 QVGTISANSDAIGAITIEAMKEKKEGVITYVEDNGENELSVYEGQGLLVHSPYFIN 204
OY 114 MEYDLAYKGLDHP-----TTHVSDIODFVVALSLEISDEGN--ITMTSEVRQFAN-V 165
DB 205 NQNNNSCELE--HPITLVDKKVSIRRLSVLE-GVAKSGRPILLIADVEGELATLV 261
OY 166 VNHIGGL-----SILDPITGVSLDVTIAIFQDTVRKEMTKVLAPARKLE 211
DB 262 VNMKGIVKVCVAKAPGFC---DRRKAMLDIAIILTKGVISSEIGKSL 308

RESULT 9
SCA4_RICMN STANDARD: PRT: 1011 AA.
ID SCA4_RICMN
AC Q9AJ82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia mongolotimonae.
OC Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=45261;
RN [1]
RP SEQUENCE FROM N.A.
RX Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT 'gene D' coding for an intracytoplasmic protein."
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF151725; AAK30683.1; -.
KW Antigen.
FT NON_TER 1
FT NON_TER 1011
SQ SEQUENCE 1011 AA; 110607 MW; 0169A06981BD5D08 CRC64;

Query Match 8.2%; Score 87.5; DB 1; Length 1011;
Best Local Similarity 22.4%; Pred. No. 19;
Matches 55; Conservative 35; Mismatches 98; Indels 57; Gaps 10;

OY 5 LIIAAVAVASADPIHDK--ITEEI--NKAID---DAIAIEOSEITIDPMKVPDHA 57
DB 402 LINTATLALSGSKQDLINYNAGITKEIDSKNKOIDIKRAAIRLNN-----KSDI 452
OY 58 FERHVGIVDFKGLAMRNIEARGLKOMKROGDANVKGEEIV-----KAHLL 104
    | : | : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gall S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwynn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001593; AAD18235.1; -
DR EMBL; AE002228; AAF38501.1; ALT_INIT.
DR EMBL; AP002545; BAA98292.1; -
DR HSSP; O9KMU6; IHQM.
DR TIGR; CP0693; -
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR002879; RNA_pol_A2.
DR Pfam; PF00623; RNA_pol_A; 1.
DR Pfam; PF01854; RNA_pol_A2; 2.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Complete proteome.
KW CONFLICT 1031 1031 A -> G (IN REF. 1).
SQ SEQUENCE 1393 AA; 154900 MW; E0734EF23C6FDE8 CRC64;

Query Match 8.6%; Score 91.5; DB 1; Length 1393;
Best Local Similarity 21.2%; Pred. No. 14;
Matches 53; Conservative 35; Mismatches 63; Indels 99; Gaps 12;

QY 36 AIAAIEGSETIDP-----MKYPDHAQKFERHGYVD 66
DB 1106 AIIIVEGQARDPGLMLARLPRGALIKTDITGSLPRVAELVARKPEDAADIAKIDGYVD 1165
QY 67 EKGELAMRNI-----EARGLKO-----MKROGDANVKG---EGIYKAHLIT--- 105
DB 1166 FKGIGKNNRIILVCDMEGMEEENLIRPLKHLIVGRDSVIGQOLTGGLVVRNHEILEIC 1225
QY 106 GVND--DIVSMEXDLAYKGLDHPHTHYISDIQDFVAL-----SLEISSEGNTMTS 156
DB 1226 GVELQKLYNVEQVRYLQV-----DINDKHIELIYQMDQKAKITDPGGTTL- 1276
QY 157 FEVROFANVNV-----HIG-----LSTIDPFGVLSVLTALAIPODYVR 195
DB 1277 -----FGEDVKKKEFEENRRTREDDGCKPAQAVPVLLGITKASLGTESTISASFODT-- 1329
QY 196 KEMTKVLAPA 205
DB 1330 ---TRVLTD 1336

```

```

RESULT 5
Y373_HUMAN
ID Y373_HUMAN STANDARD; PRT; 1539 AA.
AC O15078;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB002371; BAA20828.1; -
DR Hypothetical protein; Coiled coil.
FW DOMAIN 18 1514 COILED COIL (POTENTIAL).
SQ SEQUENCE 1539 AA; 180065 MW; D901314E981BF001 CRC64;

Query Match 8.4%; Score 89.5; DB 1; Length 1539;
Best Local Similarity 18.8%; Pred. No. 22;
Matches 29; Conservative 43; Mismatches 67; Indels 15; Gaps 3;

QY 3 KELLIAAVFAVSADPHYDKITEINKAIDDA-----IAAIEGSE--TIDPMKVPDHA 55
DB 160 KFAELITKINLDAQKQEWLRDELADSVSKAVSDARQRIILEKNEMLKYEVSLSREIS 219
DB 220 DIARQVETLANQOOSRDKREYSLRMQLLDYQAOQSDERS-----LIAKLQHNVSLQ 271
QY 56 DKFERHVGIVDFKGLAMRNIEARGLKQMKROGDANVKGEGIYKAHLILIGVHDIYSME 115
DB 116 YDLAKKLDGHPHTHYISDIQDFVALSLEISDE 149
DB 272 LSEATALGLESTISKLQKMEAYNLRLQKIDEX 305

RESULT 6
FKB4_HUMAN
ID FKB4_HUMAN STANDARD; PRT; 458 AA.
AC Q02790; Q9UCV7; Q9UCP1;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE FK506-binding protein 4 (possible peptidyl-prolyl cis-trans isomerase
DE FKB4) (EC 5.2.1.8) (PPIase) (Rotamase) (p39 protein) (HSP binding
DE immunophilin) (HBI) (FKBP52 protein) (52 kDa FK506 binding protein)
DE (FKBP59).
DE FKB4.
GN FKB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

Db 181 VLSVDVLTAFODTVRKEMTKVLAPAFKRELEKN 213
|||||
RESULT 2
ALL7_DERPT STANDARD: PRT: 215 AA.
ID ALL7_DERPT
AC P49273:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mite allergen Der p 7 precursor (Der p VII).
GN DERP7.
OS Dermatophagoides pteroyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;
OC Dermatophagoides.
OX NCBI_TaxID=6956;
RN [1]
RP SEQUENCE FROM N.A.
RA Shen H.-D., Chua K.-Y., Lin K.-L., Hsieh K.-H., Thomas W.R.;
RT "Molecular cloning of a house dust mite allergen with common antibody
binding specificities with multiple components in mite extracts.";
RL Clin. Exp. Allergy 23:934-940(1993).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE MITE GROUP 7 ALLERGEN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: U37044; AAA80264.1; .
KW Allergen; Signal.
FT SIGNAL 1
FT CHAIN 18 215 MITE ALLERGEN DER P 7.
FT CARBOHD 151 151 N-LINKED (GLCNAc...) (POTENTIAL).
SQ SEQUENCE 215 AA; 23881 MW; 63AD03DB5C8C1C0 CRC64;
Query Match 90.1%; Score 962; DB 1; Length 215;
Best Local Similarity 85.9%; Pred. No. 2.3e-70;
Matches 183; Conservative 18; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKKFLIIAAVAVAVASADPIHDKITEEINKAIDAIAIEQSETIDPKVDPHADKER 60
Db 1 MKKLLIIAAVAVASADPIHDKITEEINKAIDEAVAAIEKSETFDPKVPDHSDFER 60
QY 61 HVGIVDFKGLAMRNIEARGLKRMKGROGDANYKGEIGYKAHLIGVHDDIVSMEDLAY 120
Db 61 HIGIIDLKGLMDKRNIOVGLKRMKRVGDANYKSEGVVKAHLGVHDDVVSMEYDLAY 120
QY 121 KIGDILHPTTHVVISDIODFVVALSLEISDEGNITMTSFEVROFANVNNHIGSLIDPIFG 180
Db 121 KIGDILHPTTHVVISDIODFVVALSLEISDEGNITMTSFEVROFANVNNHIGSLIDPIFA 180
QY 181 VLSVDVLTAFODTVRKEMTKVLAPAFKRELEKN 213
Db 181 VLSVDVLTAFODTVRKEMTKVLAPAFKRELEKN 213
RESULT 3
ALL7_LEPDS STANDARD: PRT: 216 AA.
ID ALL7_LEPDS
AC Q90UG2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mite allergen Lep d 7 precursor.
OS Lepidoglyphus destructor (Storage mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

OC Acariformes; Sarcopitiformes; Astigmata; Glycyphagoidea; Glycyphagidae;
OC Lepidoglyphus.
OX NCBI_TaxID=36936;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21099292; PubMed=11168362;
RA Eriksson T.L.J., Rasool O., Huertas S., Whitley P., Crameri R.,
RA Appenzeller U., Gatvelin G., van Hage-Hamsten M.;
RT "Cloning of three new allergens from the dust mite *Lepidoglyphus*
RT destructor using phase surface display technology.";
RL Eur. J. Biochem. 268:287-294(2001).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE MITE GROUP 7 ALLERGEN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: AJ271058; CAB65963.1; .
KW Allergen; Signal.
FT SIGNAL 1
FT CHAIN 20 216 MITE ALLERGEN LEP D 7.
FT SEQUENCE 216 AA; 23916 MW; 59E71EAF05B4685E CRC64;
Query Match 23.5%; Score 250.5; DB 1; Length 216;
Best Local Similarity 29.2%; Pred. No. 2.6e-13;
Matches 63; Conservative 42; Mismatches 102; Indels 9; Gaps 5;
QY 2 MKKFLIIA-VAVAVASA--DPHDKITEEINKAIDAIAIEQSETIDPKVDPHADK 57
Db 1 MKYLAIVAVVALAGLSAAAHKRAYTD--DNMANQVNDQVKSILTRKELDPKIKGTQVP 58
QY 58 FERRHVGIDFKGLAMRNIEARGLKRMKGROGDANYKGEIGYKAHLIGVHDDIVSMEDY 117
Db 59 IDKIKGLIHIGKSAATKNAVITGLSHISRGDAKIDTDGAFATLKLG--DKNIRIKTD 116
QY 118 LAKKIGD-LHPTTHVISDIODFVVALSLEISDEGNITMTSFEVROFANVNNHIGSLID 176
Db 117 LHLDDGKIILHMKFEKGHGDIDMKIKLKLDAEGRSLDQFEIDEEVVELFIHGLPLD 176
QY 177 PIFGVSLDVLTAIFODTVRKEMTKVLAPAFKRELEK 212
Db 177 PLVDVIAIDSFVYKFFNPQARKLVTDMLKPLVVEIKK 212
RESULT 4
RPOC_CHLPN STANDARD: PRT: 1393 AA.
ID RPOC_CHLPN
AC Q92S99; Q92S99; Q9K211;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR CPN0082 OR CP0693.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 09:58:49 ; Search time 13 seconds
(without alignments)
679.574 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 1068
Sequence: 1 MMKFLLIAAVFAVAVSADPL.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1068	100.0	213	1 ALL7_DERFA	Q26456 dermatophag
2	962	90.1	215	1 ALL7_DERPT	P49273 dermatophag
3	250.5	23.5	216	1 ALL7_LEPDS	Q94192 lepidoglyph
4	91.5	8.6	1393	1 RPOC_CHLPPN	Q92999 chlamydia p
5	89.5	8.4	1539	1 Y373_HUMAN	O15078 homo sapien
6	88.5	8.3	458	1 FKBA_HUMAN	O02790 homo sapien
7	88.5	8.3	1453	1 Y373_BOVIN	O95u23 bos taurus
8	88.5	8.2	547	1 CH60_LEGPN	P26878 legionella
9	87.5	8.2	1011	1 SCA4_RICMN	Q94182 rickettsia
10	86.5	8.1	457	1 FKBA_MOUSE	P30416 mus musculu
11	85.5	8.0	903	1 YB56_METJA	Q58556 methanococ
12	85.5	8.0	1396	1 RPOC_CHLPPN	Q92K79 chlamydia m
13	85.5	8.0	1396	1 RPOC_CHLPPN	O84316 chlamydia t
14	84.5	7.9	539	1 CH60_SERRU	O86202 serratia ru
15	84.5	7.9	540	1 CH60_KLEON	O66214 klebsiella
16	84.5	7.9	540	1 CH60_KLEON	O66210 klebsiella
17	84.5	7.9	1887	1 RPB1_DROME	P04052 drosophila
18	84.5	7.9	358	1 SYFA_STROCO	O88055 streptomyce
19	84	7.9	513	1 XYIG_ECOLI	P37388 escherichia
20	83.5	7.8	540	1 CH60_ENTGE	O66194 enterobacte
21	83.5	7.8	550	1 CH60_YEREN	P46210 yerushla en
22	83	7.8	244	1 RS4E_METJA	P54039 methanococ
23	83	7.8	465	1 APEA_CLOAB	Q97K30 clostridium
24	83	7.8	633	1 DNAK_AGRFS	P56019 agrobacteri
25	83	7.8	1060	1 DP3A_LACIA	Q96170 lactococcus
26	82.5	7.7	539	1 CH60_ENTAM	O66196 enterobacte
27	82.5	7.7	540	1 CH60_KLEPL	O66212 klebsiella
28	82.5	7.7	540	1 CH60_SERPI	O66204 serratia fi
29	82.5	7.7	2483	1 COA2_HUMAN	O00763 homo sapien
30	82	7.7	461	1 DUDH_CHLPPN	O92773 chlamydia p
31	82	7.7	2104	1 MTS3_SCHPO	O14157 schizosacch
32	81.5	7.6	541	1 CH60_PANAN	O66218 pantoea ana
33	81.5	7.6	547	1 CH60_HAEDU	P31294 haemophilus

34	81.5	7.6	745	1 ATCS_SYNX3	P73241 synechocyst
35	81	7.6	244	1 RS4E_METJA	P14023 methanococ
36	81	7.6	798	1 UNR_HUMAN	Q75534 homo sapien
37	80.5	7.5	159	1 BRR_SERMA	O68935 serratia ma
38	80.5	7.5	194	1 VATE_SULFO	O971B8 sulfolobus
39	80.5	7.5	539	1 CH60_ENTAG	O66200 enterobacte
40	80.5	7.5	539	1 CH60_ENTAS	O66190 enterobacte
41	80.5	7.5	540	1 CH60_ENTAE	O66198 enterobacte
42	80.5	7.5	540	1 CH60_ERMAP	O66222 erwiniola aph
43	80.5	7.5	541	1 CH60_ERHME	O66216 erwiniola her
44	80.5	7.5	1010	1 SCA4_RICPA	Q94J75 rickettsia
45	80	7.5	160	1 TPC2_CAEBL	Q09665 caenorhabdi

ALIGNMENTS

RESULT 1	ID	ALL7_DERFA	STANDARD:	PRT:	213 AA.
AC	Q26456:				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Mite allergen Der f 7 precursor (Der f VII).				
GN	DERP7.				
OS	Dermatophagoides farinae (House-dust mite).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;				
OC	Acariiformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;				
OC	Dermatophagoides.				
OX	NCBI_TaxID=6954;				
RN	[1]				
RP	SEQUENCE FROM N. A.				
RA	MEDLINE=96120794; PubMed=8556554;				
RT	Shen H.-D., Chua K.-Y., Lin W.-L., Hsieh K.-H., Thomas W.R.;				
RT	"Molecular cloning and immunological characterization of the house				
RT	dust mite allergen Der f 7."				
CL	Clin. Exp. Allergy 25:1000-1006(1995).				
RC	- SUBCELLULAR LOCATION: Secreted.				
CC	- SIMILARITY: BELONGS TO THE MITE GROUP 7 ALLERGEN FAMILY.				
CC	-				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-				
DR	EMBL: S80655; AAB35977.1; -.				
KM	Allergen; Signal.				
FT	SIGNAL	1	17	POTENTIAL.	
FT	CHAIN	18	213	MITE ALLERGEN DER F 7.	
FT	CARBOHYD	151	151	N-LINKED (GLCNAc. . .) (POTENTIAL).	
FT	SEQUENCE	213 AA: 23627 MW; 3CF1F529107B7808 CRC64;			
QY	Query Match	100.0%:	Score 1068:	DB 1:	Length 213:
QY	Best Local Similarity	100.0%:	Pred. No. 7.2e-79:		
QY	Matches 213:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY	1 MMKFLLIAAVFAVAVSADPIHYDKTTEETINKAIDDAIAIEQSETIDPMKVPDHADEER 60				
DB	1 MMKFLLIAAVFAVAVSADPIHYDKTTEETINKAIDDAIAIEQSETIDPMKVPDHADEER 60				
DB	61 HVGIVDFGELAMRNIEARGLKQMRQGDANYKGEQIVKAHLILGVHDDIVSMEDLAY 120				
QY	61 HVGIVDFGELAMRNIEARGLKQMRQGDANYKGEQIVKAHLILGVHDDIVSMEDLAY 120				
QY	121 KIGDILPHTTHVSDIDODPVVALSLSEISDEGNTTHSFEYKQANVYVNHIGSLIDPIFG 180				
DB	121 KIGDILPHTTHVSDIDODPVVALSLSEISDEGNTTHSFEYKQANVYVNHIGSLIDPIFG 180				
QY	181 VLSVLTATIFQDTVRKEMTKVLAPAFKRELEKN 213				

Db 105 DRGRKLRKPTENPDVKLCKIKNKRYIKG-----GHIOQLNLDHGRNIVIKVSDPTKAE 158
QY 117 DLAYKIGDLHPPTHVVISDIDQDFVALSLSDGNITMTSEVROFANVY--NHIGLSI 174
Db 159 DV-YKRGD-----TLISIPEDIRKAHFPEVGKLAIVYGKHAVGDFAK 201
QY 175 LDPI--FGVLSVLT 187
Db 202 IVEIERRGIVPIVY 216

RESULTS42

F97034
aspartyl aminopeptidase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97034
R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97034
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <KUR>
A:Cross-references: GB:AE001437; PIDN:AK79065.1; PID:q15024007; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1091

Query Match 7.8%; Score 83; DB 2; Length 465;
Best Local Similarity 28.4%; Pred. No. 60;
Matches 42; Conservative 27; Mismatches 49; Indels 30; Gaps 10;

QY 74 RNIE---ARGLKQKROGD---ANVR-----GEGIVKAHLIGVHDI--VSMEX 116
Db 58 RNIEDILAKG--ETLKEGDKVYANNRKGKGLIMFLGKPLGTGFKILGAHIDSPRLDKQ 115
QY 117 DLAYKIGDLHP--TTHVISDIDQFV--VALSLSDGNITMTSEVROFANVNHIGLSI 174
Db 116 NPLYEDTDLAMLETHYGGIKTYQWTLPLAT--HGVI-----VKKDGTIVNVCGEDD 167
QY 175 LDPIFGVLSVLTAFIDFQDVRKEMTKVL 202
Db 168 NDPEVGV-SDILVHLASQQLKAKSKYI 194

RESULT 43

I39585
dnaK-type molecular chaperone dnaK - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 20-Aug-1999
C:Accession: I39585
R:Segal, G.; Ron, E.Z.
J. Bacteriol. 177, 5952-5958, 1995
A:Title: The dnaK operon of Agrobacterium tumefaciens: transcriptional analysis and ex
A:Reference number: I39585; MUID:96011387; PMID:7592349
A:Accession: I39585
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-633 <RES>
A:Cross-references: EMBL:X87113; NID:q1027503; PIDN:CAA60592.1; PID:q1027504
C:Genetics:
A:Gene: dnaK
C:Function:
A:Description: Involved in protein folding and assembling/disassembling of protein comp
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 7.8%; Score 83; DB 2; Length 633;
Best Local Similarity 28.3%; Pred. No. 87;

Matches 28; Conservative 16; Mismatches 27; Indels 28; Gaps 4;
QY 23 DKITEEINKAIDDAIA----AIEQSETIDPMKVPDHADEKFERHVGIVDFKGLAMRNEA 78
Db 550 DKVSEIDRKAIEDAIASLKTAVEAE-----PDADDIOAKTQTLMEVSMKLGQATYEA 602
QY 79 RGLKQKROGDANVKEGEGIVKAHLIGVHDIYSMEYD 117
Db 603 ----QQAEGADASAEK-----DDVADADYE 624

RESULT 44

E97373
dnaJ protein (heat shock protein 70) (hsr70) [imported] - Agrobacterium tumefaciens (C
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: E97373
R:Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: E97373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-633 <KUR>
A:Cross-references: GB:AE007869; PIDN:AK85942.1; PID:q15154995; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_195
A:Map position: circular chromosome
C:Superfamily: heat shock protein 70

Query Match 7.8%; Score 83; DB 2; Length 633;
Best Local Similarity 28.3%; Pred. No. 87;
Matches 28; Conservative 16; Mismatches 27; Indels 28; Gaps 4;

QY 23 DKITEEINKAIDDAIA----AIEQSETIDPMKVPDHADEKFERHVGIVDFKGLAMRNEA 78
Db 550 DKVSEIDRKAIEDAIASLKTAVEAE-----PDADDIOAKTQTLMEVSMKLGQATYEA 602
QY 79 RGLKQKROGDANVKEGEGIVKAHLIGVHDIYSMEYD 117
Db 603 ----QQAEGADASAEK-----DDVADADYE 624

RESULT 45

AE2591
DNAK protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AE2591
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Giller, W.; Grant, C.; Genthner, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE2591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-633 <KUR>
A:Cross-references: GB:AE008668; PIDN:AA141147.1; PID:q17738443; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: dnaK
A:Map position: circular chromosome
C:Superfamily: heat shock protein 70

Query Match 7.8%; Score 83; DB 2; Length 633;
Best Local Similarity 28.3%; Pred. No. 87;
Matches 28; Conservative 16; Mismatches 27; Indels 28; Gaps 4;


```

Db 120 TLROCKLLAQVSLSPPTRRVGDGLGGQQLVEIKALKNQVRLILDEPPASLTREQETS 179
QY 189 IFQDTR 195
| | : |
Db 180 ILDDIR 186

RESULT 35
T17436
ATP-binding protein ybtQ - Yersinia pestis
C:Species: Yersinia pestis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C:Accession: T17436
R:Bertherton, J.D.; Bertolino, V.J.; Perry, R.D.
Mol. Microbiol. 32, 289-299, 1999
A>Title: YbtP and YbtQ: two ABC transporters required for iron uptake in Yersinia pestis
A:Reference number: 218782; PMID:99248409; PMID:10231486
A:Accession: T17436
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-600 <FEF>
A:Cross-references: EMBL:AF091251; NID:g3818595; PID:g3818601; PIDN:AC69584.1
R:Buchrieser, C.; Ruskick, C.; Couve, E.; Frangeul, L.; Billault, A.; Kunst, F.; Carnieja
submitted to the EMBL Data Library, October 1998
A:Description: DNA sequence of the 102 kbases unstable region of Yersinia pestis.
A:Reference number: 224348
A:Accession: T47044
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-600 <BUC>
A:Cross-references: EMBL:AL031866; PIDN:CAA21387.1
A:Experimental source: strain 6/69
C:Genetics:
A:Gene: ybtQ
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 7.9%; Score 84; DB 2; Length 600;
Best Local Similarity 21.6%; Pred. No. 68;
Matches 42; Conservative 27; Mismatches 65; Indels 60; Gaps 6;

QY 6 LIAAFAFAVASADP----IHYDKITEELINKAIDDAIAIEQSETIDPKVDPHADKFERH 61
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 287 LIAAFAFAFAVASADP----IHYDKITEELINKAIDDAIAIEQSETIDPKVDPHADKFERH 61
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 62 VGIVDFKGLAMRNIEARGELKQMKRGDANVKGEGCIKAKHLLIGVHDDIVSMEDLAYK 121
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 343 -----DIRDNVSYSR-----YE 354
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 122 LQDLHPHTHYVLSIDIDFVVALSLSEISDEGNITMTSEFYKQFANVYN---HIGGSLDPT 178
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 355 EGDGALHNVSLTFPPAASMSALVGASGAGKTTVTKLMR-YADPOGQISIGGVDIRRLT 413
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 179 FGVLSDEVLTALFOD 192
| : : : : | | |
Db 414 PEQLNSLISVVFQD 427

RESULT 36
AF0233
Inner membrane ABC-transporter YbtQ ybtQ [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C:Accession: AF0233
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tilball, R.W.; Holden, M.T.G.; Prentice, M.B
deleno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0233
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <KUR>

```

A:Cross-references: GB:AL590842; PIDN:CAC90730.1; PID:g15979933; GSPDB:GN00175
C:Genetics:
A:Gene: yf9c
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 7.9%; Score 84; DB 2; Length 600;
Best Local Similarity 21.6%; Pred. No. 68;
Matches 42; Conservative 27; Mismatches 65; Indels 60; Gaps 6;

QY 6 LIAAFAVAVSADP---IHVDKITEEINKAIDAIAIEQSETIDPKKVPDHADKFERH 61
DB LIAAFAAMIRFAEPFAMFISYTSVE---LIASALQRIEFPMAIAPLPAVEOSEMPERY 342
QY 62 VGIYDFKELAMRNIEAAGLQKMKRGDANKGEGYKKAHLILGVHDDIYSMEYDLAK 121
DB 343 -----DIRPDNYSYR-----YE 354
QY 122 LGDLHPTRHVISDIQDFVVALSLSEISDEGNITMTSFEEYROPANVYN---HIGGLSILDP 178
DB 355 EGDGHALNHVSLTFPPAASMSALVGASGAKTYYTKLMR-YADPQQGISIGVDIRLRT 413
QY 179 FGVLSDVLTALFOD 192
DB 414 PEQLNLSISVVFQD 427

RESULT 37

G86696
cation-transporting ATPase yf9c [Imported] - Lactococcus lactis subsp. lactis (strain
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
R:Accession: G86696
R:Boletijn, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarre, K.; Weissbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: G86696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-775 <STO>
A:Cross-references: GB:AA005176; PID:g12723465; PIDN:AAK04673.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yf9c

Query Match 7.9%; Score 84; DB 2; Length 775;
Best Local Similarity 19.8%; Pred. No. 93;
Matches 50; Conservative 37; Mismatches 103; Indels 62; Gaps 6;

QY 22 YDKITEEINKAIDAIAIEQSETIDPKKVPDHADKFERHVGIDFKGL-----AMRN 75
DB 397 YEKVSEIINHLEEGFRLVLATGKE--KIYDQLMGAYVALGYVVLNPIRENAKSTFNY 454
QY 76 IEARGLQKMKRGD-----ANKGEGYKKAHL----- 104
DB 455 FAEGGVNIAKVISGDPQTVSAVAKKAGITGAEKRFIDANLLTKEDLDQAVSYTFGRVT 514
QY 105 -----IGVHDDIYSMEYD-----LAVKLGDLHPTRHVISDIQDFVVALSLSEISDE 149
DB 515 PDQKRLVQALKKRKHVTAMGCGVNDILAMKSADCSIAMASGSDPATQVAQVVLDSDF 574
QY 150 GNITFTSEVQAFANVYVHHIGLSILDPIFGVLSVLTALFQDYPRKMKTK-----V 201
DB 575 GHTQVTVTEGRVANNVORSATILFLVKNLFSTILAIISAFVFTYPILOASOLISLFTI 634
QY 202 LAPAFKRELEKN 213
DB 635 GIPGFLSLEEN 646

RESULT 38

G97003
beta-glucosidase family protein [Imported] - Clostridium acetobutylicum

```

Oy      110 D-----IVSMNEYDLAKKGLDHPTT---HVISIDQFPVALSLEIDEGNI-----T 153
Db      806 DYGPSRGFEVNSY-----LAGLTPSEFYFHAMGG-REGLLDTAVKTAETGYIORLLIKA 859
Oy      154 MTSFEVROFANVNVNHIGLSIL---DPFGVLSL----- 184
Db      860 MESVAVNVNDGIVTRNSVGQLQIRVGEDGLGELVFEFGNMPYKLSNKSFEKRFKRWDSNE 919
Oy      185 -VLTAIFODTYVRKEMT 199
Db      920 RLMKKVFPTDDVIKEMT 935

RESULT 32
A72335
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72335
R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: A72335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-314 <ARN>
A:Cross-references: GB:AE001747; GB:AE000512; NID:g4981304; PIDN:ADJ35882.1; PID:g498132
A:Experimental source: strain MSB8
A:Genetics:
A:Gene: TM0800

```

```

Query Match          7.9%, Score 84; DB 2; Length 314;
Best Local Similarity 23.5%, Pred. No. 31;
Matches 48; Conservative 32; Mismatches 76; Indels 48; Gaps 10;

QY      23 DKITEIKKAIIDALIAA--IEGSEIIDPMKVPDHAHKREERY-----GIWPKGEELA 72
           | : : : || : : | : : : : : | : : | : : | : : |
DB     122 DSLARVMEAGADAVIAEGMESGHIGVTTFVLVNKRSRASNIPVIAGGIADGRGMAA 181
           | : : : : : : : : : : : : : : : : : : : : : : : : :
QY      73 MRNIEARGLKQMR-----OGDANKGEGGIYKAHLILGVHDDIVSMGYDLAYRKLGLHP 127
           | : : : : : : : : : : : : : : : : : : : : : : : : :
DB     132 AALCAEAQVMQTRRVASVESDVHPYKEKIYKA----SIRDTVYT-----GAKLG--HP 230
           | : : : : : : : : : : : : : : : : : : : : : : : : :
QY     128 T-----THVISDIODEFVALLEISID-----EGNITMTSPHEQPFANVVHHIIGL 172
           | : : : : : : : : : : : : : : : : : : : : : : : : :
DB     231 ARVLTPPRARKIQEHEFNPMQAEMLVGSLRRAYVEDDLERGSIMWGQSAGLIDET--- 287
           | : : : : : : : : : : : : : : : : : : : : : : : : :
QY     173 SLIDPIFGVLSVDVLTAFIDFTVRK 196
           | : : : : : : : : : : : : : : : : : : : : : : : : :
DB     288 ---RPVKQIIEDILKE-FKEIVEK 307

RESULT 33
GB99774
phosphopentomutase [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: GB99774
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogucuma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: AB9758; MWID:21311952; PMID:11418146
A:Accession: GB99774
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-392 <KUN>
A:Cross-references: GB:BA000018; PID:d13700055; PION:BAB41354.1; GSPDB:GN00149
```

A;Experimental source: strain N315
C;Genetics:
A;Gene: drm
C;Superfamily: phosphopentomutase

Query Match	7.9%	Score 84	DB 2	Length 392
Best Local Similarity	24.1%	Pred. NO. 41		
Matches 45; Conservative	35;	Mismatches	85;	Indels 22;
				Gaps 10;

```

QY 14 AVSAPRIHYDKTTE-EINKAIDALDAALAEQSEITIDPMKVRPNADKFEHNAGIVDFKELA 72
Db 65 AAEQPEAYVYTKISEASVSKDITMTGHWELMTGLINMQPEFVYVYNGFPREELIQIEEMTGRKV 124
QY 73 MRNIEARGLOKMRQGDANVNGEEGIVAKH---LLIGVHDDIYSME--YDLAKTLDL- 125
Db 125 VANKRVAISGTQIIDEENGEOHMTGDLIVYTSADPVYQIATNHHEDTIPLELHYTICRKAVELT 168
QY 126 HPTTVHISDIDQEVVALSLEISDEGNITMTS---FEVYROFA-NVYNIH--GGLSIDLPI 178
Db 185 KDPKYLIGRI---IARPY-VGEPGNFTPRTSNRHDYALKPEGKVTYLDHLKDGQYDI--A 237
QY 179 FGVLSDV 185
Db 238 IGRINDI 244

```

RESULT 34
S47788
d-xylose transport ATP-binding protein xylG - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 27-Jan-1995 #sequence_revisions 27-Jan-1995 #text_change 01-Mar-2002
C:Accession: S47788; A65156
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994

A:Reference number: 347000
A:Accession: S47788
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-513 <PIII>
A:Cross-references: EMBL:U00039, NID:9466582, PIDN:AAB18544.1, PID:9466705
R:Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shaoh, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A65156
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-513 <BLAT>
A:Cross-references: GB:AE000434, GB:U000096, NID:g1789989, PIDN:AAC76591.1, PID:g178999
A:Experimental source: Strain K-12, substrain MG1655
C:Genes: xylG
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP, nucleotide binding, P-loop
F:20-218/Domain: ATP-binding cassette homology <ABC1>
F:37-44/Region: nucleotide-binding motif A (P-loop)
F:277-481/Domain: ATP-binding cassette homology <ABC2>

```

Query Match 7.98; Score 84; DB 2; Length 513;
Best Local Similarity 23.58; Pred. No. 56;
Matches 44; Conservative 32; Mismatches 71; Indels 40; Gaps 7;

QY 31 KAIDDAIAIESEFID-----PMVPPHADKFERHVGIVDKEGLAMRNIEAR 79
      |||: : : : : ||| : : : : ||| : : : : |||
Db 18 KAIIDVCLRLNGELVSLDGENGSKSLTMVLCIYFHSGEGLIFAGE-----ELDAS 73

QY 80 GLKQKRGQDANVCKEGIVK-----AHLIG--VHDDIVSMEDLAYKIGDLHPPTHV 131
      : : : : : : : ||| : : : : : : : ||| : : : : : : : |||
Db 74 HIRDERKGIALLHQLLALVKEVLLENIFLGNEIITHNGI--MYDYL-----M 119

QY 132 ISDIDDFVALSLFLEISDE---GNITMTSEFEVROFANVNVNHIGGSIIDPIFGVLSDVLT 188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


[illegible][illegible]

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: F81686
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: F81686
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1396 <TEXT>
A:Cross-references: GB:AE002328; GB:AE002160; NID:97190627; PTDN:AAF39420.1; PID:9719
A:Experimental source: strain Nigg (Mopn)
A:Genetics:
A:Gene: TC0588
C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain

Query Match 8.0%; Score 85.5; DB 2; Length 1396;
Best Local Similarity 23.6%; Pred. No. 1.5e+02;
Matches 49; Conservative 34; Mismatches 66; Indels 59; Gaps 12;

[illegible]

Db 193 GE 194

RESULT 22

B82048
Chaperonin, 60 Kd chain VC2664 [imported] - Vibrio cholerae (strain N16961 serogroup O1)C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001C:Accession: B82048
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;Chaidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
L. R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: B82048

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-544 <HEI>

A:Cross-references: GB:AE004332; GB:AE003852; NID:g9657252; PIDN:AAF95805.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2664

A:Map position: 1

C:Superfamily: chaperonin groEL

Query Match 8.1%; Score 86.5; DB 2; Length 544;
Best Local Similarity 24.4%; Pred. No. 39;
Matches 41; Conservative 21; Mismatches 69; Indels 37; Gaps 5;QY 6 LIAAFAFVAVSA-----DPHYDKITEINKAIDAIAIEOSEITIDPMKV 51
DB 372 LAGGAVAVIVGATEVEEMKEKKRDVEDALHATRAAVEEGVAGGALLRAASKLSSTLV- 430
QY 52 PPHADKFERHVGIVDFKGLAMRNIEARGLKQKROGA-----NKVEEGIVKAHLII 105
DB 431 ---GNNEENNGI-----RVALLRAMEAPLRQIVKNAGDESVANNVRAGEVGYNAAT 482QY 106 GVHDDIVSMEDYLAAYKLGDLHPTHTHISIDIOFVVALSLEIDEGNIT 153
DB 483 GAYGDMEI-----GILDPTKVTRSALQFPAASYAGLMTTEAMIT 522

RESULT 23

DB3482
Probable transcription regulator PA1312 [imported] - Pseudomonas aeruginosa (strain PA01)C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000C:Accession: DB3482
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lam,
.. Lody, S.; Olson, M.V.Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: DB3482

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-303 <STO>

A:Cross-references: GB:AE004560; GB:AE004091; NID:g9947240; PIDN:AAG04701.1; GSPDB:GN001

A:Experimental source: strain PA01
C:Genetics:A:Gene: PA1312
C:Superfamily: conserved hypothetical protein H11364Query Match 8.1%; Score 86; DB 2; Length 303;
Best Local Similarity 22.6%; Pred. No. 21;
Matches 28; Conservative 19; Mismatches 39; Indels 38; Gaps 4;QY 24 KITEENKAIIDAIAIEGSETIDPMKVPDHDKFERHVGIVDFKGLAMRNIEARGLQ 83
DB 71 KYIDEARLAIDARA-----EHG-----ELKSLRVYTTQYGLNQ 106QY 84 MKRGDANKGEGIVKAHLIGV-----HDDIVSMEDYLAAYKLGDLHPTHTHISIDIQ 136
DB 107 L-----VPALQAFARLHPALVOLSTSLHADLIGERFVAIRLGRLEDSHTHAVOLA 159QY 137 DFVV 140
DB 160 SFEV 163

RESULT 24

E70710
Hypothetical protein Rv1482c - Mycobacterium tuberculosis (strain H37RV)C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: E70710

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd,Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: E70710

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-339 <COL>

A:Cross-references: GB:E79701; GB:AL123456; NID:g3261635; PIDN:CAB02032.1; PID:g15242

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv1482c

C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1073

Query Match 8.1%; Score 86; DB 2; Length 339;
Best Local Similarity 28.3%; Pred. No. 24;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;QY 33 IDDAIAIEOSEITIDPMKVPDHDKFERHVGIVDFKGLAMRNIEARGLKQKR----- 86
DB 187 LBPAAVAADVALIQADVLKADVEPLIERYGRGRKMAARAALVDGAGQSPKRWLRLL 246QY 87 -----QGANVKGEGIVKAHLIGVHDDIVSMEDYLAAYKLGDLHPT--HVISI 135
DB 247 LIRAGFPFPQTOIAVRENEMWMAEHLDMGMODIKAAAEYD-----GDHHLTSRYHYRKDI 301

RESULT 25

C64444
Cell division control protein CDC48 homolog - Methanococcus jannaschiiC:Species: Methanococcus jannaschii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: C64444

R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weissstock, K.G.; Merrick, J.M.; Glodek,
ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999; PMID:8688087

A:Accession: C64444

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-903 <BUL>
A:Cross-references: GB:U67557; GB:L77117; NID:g1591777; PIDN:AAB99153.1; PID:g1591785C:Genetics:
A:Map position: REV1097124-1094413
C:Superfamily: transitional endoplasmic reticulum ATPase; FtsH/SEC18/CDC48-type ATP-b
F:193-400/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VARI>
F:220-227/Region: nucleotide-binding motif A (P-loop)
F:466-674/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VARI>
F:493-500/Region: nucleotide-binding motif A (P-loop)

Query Match 8.0%; Score 85.5; DB 1; Length 903;

D0 67 POAKAEACKRPMAGCAANKALAAQEELEAERDARLVEEAV-----DVTLP 114

Oy 116 YDLAKKGLDHPTRHVIDIODEFVALSLETSD 148
| | | | |
Db 115 HD-RVPAGARRHPTLTLSERIEDIFVAMGIVEAE 146

RESULT 6

B86d239

protein T10024.19 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: B86d239

R:rtetologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,
ansen, N.F.; Hughes, B.; Hutzar, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Malli, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Saldegg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:1130712

A:Accession: B86d239

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-697 <STO>

A:Cross-references: GB:AEO05172; MID:g50591550; PIDN:ADJ39579.1; GSPOB:GN00141

C:Genetics:

A:Map position: 1

RESULT 7
B90601
hypothetical protein MYPU_7140 [imported] - Mycoplasma pulmonis (strain UAB C1P)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: B90601
R:Chambud, I.; Heiligg, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.:
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: B90601
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-979 <KUR>
A:Cross-References: GB:AL445566; PID:g14090129; PIDN:CA01387.1; GSPDB:GNO0153

```

A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU_7140
A:Genetic code: SGC3

Query Match          8.6%; Score 91.5; DB 2; Length 979;
Best Local Similarity 21.2%; Pred. No. 33;
Matches 46; Conservative 45; Mismatches 93; Indels 33; Gaps 9;

OY      3 KFLIATAAFVAVASDPRHYDKIT-----EINKAIDATAIAIQSETIDPMK 50
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      616 EFSVAALVYERKLSIDKOSFDSLTAAYOTISAKDLGESSSSALDQINELIQAKEYI--LQ 673

OY      51 VPDHADKEFERHNGIYDFGELAMRNIPEARGL-KOMKRGDANVKGEEGIVRAHLIGVD 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      674 AQIRAK-----ESLISLMTDFSXYLLENKNDLESLRIINDEFSKKITPKKNAEYELDINQ 729

OY      110 DIVSMEYDLAYKLG-DLHPTTHVISDIODFYVAASLEISDEGNITMTSEFYQAFANVNH 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      730 KRLSEFYMMWTEMAKDNFSLSIDITEAEFPVGRIKEETIKDGININIDS--IKSYEEISER 787

OY      169 IGGSLIDPIFGVLSDVLTALFODPVR--KEMTKVL 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      788 YDGLKATE-----SEIYDA-KKQIESAIKEMDKII 816

```

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
36 A1AAE0SETDP	8.6%; Score 91.5; DB 2; Length 1393;	53;	35;	63;	99;	12;
1106 A1ISVE0ORVDPGMLRLPRGAIKTKDTGTGRLPRVAELVEARRPEDDAADIAKIDGVVD						
67 FKGEIAMANI	1166 FKGIQKNNRILVCDCEMTGMEBEENHILPLTKNHLIVRGGSVYKGGQULPTGLVVPHPHILEIC					
106 GVHD-DIVSMEXYDLAVYLGDLHPTTHAVISDIQDFVAL	1225					
1226 GVRELOKTYLVNEVOEYRLOGV	156					
157 FEVRQFANVVN	1276					
1277 FGDVDNKKFEENNRRTEDDGRPAQAVPVLLGITKASIGTESFISASFQDT	195					
196 KEMTKVLAP	1329					
1330 TRVLFDA						

66 -DEKELAMNIEARGL-----KOMKROGDANVKGEGIVAKHLLIGVNDIYSME 115
| | | | | : | | : | :
QY

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:00:50 : Search time 21 Seconds
(without alignments)
975.078 Million cell updates/sec

Title: US-10-024-955-7
1068

Perfect score: 1 MMKFLIAAFAVAVSADPI.....VRKEMTKVLAFAKRELEKN 213
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	10.2	568	2 T49962	hypothetical prote
2	97.5	9.1	776	2 E90591	lipoprotein (impor
3	93.5	8.8	449	2 F84241	hypothetical prote
4	93.5	8.8	628	2 G75154	tungsten-containin
5	92.5	8.7	373	2 T36830	probable phenylala
6	92.5	8.7	697	2 B86239	protein T10024.19
7	91.5	8.6	979	2 B90601	hypothetical prote
8	91.5	8.6	1393	2 E72122	RNA polymerase bet
9	91.5	8.6	1393	2 B86501	DNA polymerase bet
10	91.5	8.6	1397	2 E81548	DNA-directed RNA p
11	90.5	8.5	547	2 F82783	60kDa chaperonin x
12	90.5	8.5	681	2 T15590	hypothetical prote
13	90	8.4	1558	2 B71603	RESA-H3 antigen pr
14	88.5	8.3	459	2 A46372	immunophilin FKBP5
15	88.5	8.3	2137	2 T05244	hypothetical prote
16	88	8.2	436	2 B95199	phosphoglycerate d
17	88	8.2	436	2 H98065	phosphoglycerate d
18	87	8.1	372	2 T32933	F9K20.3 (imported)
19	87	8.1	399	2 T32933	hypothetical prote
20	87	8.1	1640	2 D86798	prophage p13 prote
21	86.5	8.1	458	2 JN0873	immunophilin p59 -
22	86.5	8.1	544	2 B82048	chaperonin, 60 kd
23	86.5	8.1	303	2 D83482	probable transcrip
24	86	8.1	339	2 E70710	hypothetical prote
25	85.5	8.0	903	1 C64444	cell division cont
26	85.5	8.0	1396	2 G71529	DNA-directed RNA p
27	85.5	8.0	1396	2 F81686	DNA-directed RNA p
28	85.5	8.0	1441	2 T00335	hypothetical prote
29	85	8.0	292	2 G75084	hypothetical prote

30	85	8.0	878	2 F86795	cation-transporlin
31	84.5	7.9	1896	1 RNEF2L	DNA-directed RNA p
32	84	7.9	314	2 A72335	conserved hypothet
33	84	7.9	392	2 G89774	phosphopentomutase
34	84	7.9	513	2 S47788	d-xylose transport
35	84	7.9	600	2 T17436	ATP-binding protei
36	84	7.9	600	2 AF0233	inner membrane ABC
37	84	7.9	775	2 G86636	cation-transporlin
38	83.5	7.8	474	2 G97033	beta-glucosidase f
39	83.5	7.8	548	2 AC0043	60 kDa chaperonin
40	83.5	7.8	550	2 S52901	heat shock protein
41	83	7.8	244	2 D64358	ribosomal protein
42	83	7.8	465	2 F97034	aspartyl aminopept
43	83	7.8	633	2 I39585	dark-type molecula
44	83	7.8	633	2 P97373	dnaj protein (heat
45	83	7.8	633	2 AE2591	DNAA Protein (limpo

ALIGNMENTS

```

RESULT 1
T49962
hypothetical protein F8M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49962
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: 22493
A:Accession: T49962
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-568 <BEV>
A:Cross-references: EMBL:ALJ33993; GSPDB:GN00063; ATSP:F8M21.160
A:Experimental source: cultivar Columbia; BAC clone F8M21
C:Genetics:
A:Gene: ATSP:F8M21.160
A:Map position: 5
A:Introns: 200/3; 337/3; 544/3

Query Match          10.2%  Score 109; DB 2; Length 568;
Best Local Similarity 26.7%; Pred. No. 0.74;
Matches 46; Conservative 36; Mismatches 64; Indels 26; Gaps 8;

QY 25 TTEINKAIDDAIAIEQSETIDPMK--VPDIADFERHVGIVDRKGEIAMN-----IE 77
      || : ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 331 ITISAREVEPEDA-----YSPTEAVWRLOPKCSDKVERDSGLVFTTRLVPSRIGCIL 385
QY 78 ARG---LKKMKROGDANVK--GEEGIKVA---HLLIGVHDDIVSMEDYLAVKLGIDLPPT 128
      :| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 386 GKGGAITTEMKTKANINILKENLPKVASDDDEVOVNFVALLKFSLOGLLSLRKF 445
QY 129 TVHISDIDQFVVALS--LEISDEGNITMTSFEVKOFANVNHIGGISLIDPT 178
      | : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 TMSVNSNSNFVDVQISGLDPAKEALITQITS---RLRANFDEGAVSALMPV 494

RESULT 2
E90591
lipoprotein (imported) - Mycoplasma pulmonis (strain UAB CTRP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: E90591
R:Chamblaud, I.; Helliou, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A>Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A:Reference number: A95512; MUID:21267165; PMID:11353084
A:Accession: E90591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <KUR>
A:Cross-references: GR:AL445566; PID:gl4090052; PIDN:CAC13810.1; GSPDB:GN00153

```


Db 2346 IQNVNKLADKLK-EIE 2361

RESULT 50

US-09-808-880-2

; Sequence 2, Application US/09808880
; Publication No. US20030027287A1

GENERAL INFORMATION:

; APPLICANT: Betlach, Mary C.
; APPLICANT: Shab, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert

; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880

; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517

; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254

; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100

; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4150

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant

; OTHER INFORMATION: Oleandolide PKS

US-09-808-880-2

Query Match 6.8%; Score 73; DB 9; Length 4150;
Best Local Similarity 19.5%; Pred. No. 4.7e+02;

Matches 44; Conservative 40; Mismatches 74; Indels 68; Gaps 11;

QY 9 AAVAFYAV-----SADPIHYDKITEINKAIDDAIAIEQSEIDPMKVP-----DHADK 57
DB 11 STAIYGIACRLPGSANTPOEFMWRLADSNADALDEPRAGRPFGSLSPRAPRGFLDSIDT 70
QY 58 FE-----RHVGIVDEKGEIAM-----RNIEARGIKOMKROGD-----89
DB 71 FDADDFNISPREAGVLDPOQRLALEIGWEALDEDAGIVPRHLRGTRTSVPMGAMMDYAH 130
QY 90 ANVKGEGEIVKAHLIGVDDIVSMEDLAYKLGDLHTTHVYISDIQDFVALSLEISD- 148
DB 131 AHAKGEAALTR-HSLGTGRMIANR--LSYALGLOGPSLTVDTGQSSSLAAVHMACESL 187
QY 149 --EGNITWTSFEVROFANVNVNIGLS-IIDPI-----FGVLS 183
DB 168 ARGESDIAL-----VGGVNLVLDPA GTTGVERRGALS 219

Search completed: February 21, 2003, 10:14:19
Job time : 41 secs


```

: LENGTH: 1993
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-098-979-2

Query Match
Best Local Similarity 23.1%, Score 73.5; DB 9; Length 1993;
Matches 45; Conservative 33; Mismatches 82; Indels 35; Gaps

QY 13 VAVSADPIHYDKITEINKAIDDAIAIEQSEITDPMKVPYDHADKFERHHVGIYDFKGLA 72
Db 1518 VALSGDHALESQIRQLGKALDDSRFQIQOTENIRSKPTPTGPELDTSYKGNKLLGSEC- 1576
QY 73 MRNTEARLKKMKRGCDANVKEE---GIYAH---LLIGVH-----DIYSMEID 117
Db 1577 -----SSSIDSVKRL-EHKLKEEESLDPGFVNLHSTETQTAQVDIRWELLQAAALSKELR 1630
QY 118 LAYKL-----GDDHPPTHVHISDIDQDFVALS-LEIS-DEGNITWTFPEVQFANVNH 168
Db 1631 MKQNIQKQKQRFSDJNSTWAMLGDTBEELQJLRLELSTDIOTIELQIKKLEKQAVDH 1690
QY 169 ---TGLSLDPIF 179
Db 1691 RKATILSLNCSPEF 1705

RESULT 46
US-09-815-242-10232
: Sequence 10232, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10232
: LENGTH: 361
: TYPE: PRT
: ORGANISM: Escherichia coli
US-09-815-242-10232

Query Match
Best Local Similarity 20.8%; Score 73; DB 10; Length 361;
Matches 38; Conservative 28; Mismatches 45; Indels 72; Gaps

QY 52 PDHAD-KFERHHVGIYDFG-----ELAMR-----NIFARG----- 80
Db 104 PCHADVYTEQKYGILDRGGRSSARERFAMVAAAGAIKRYLAERFGIETRGCTLQMGDI 163

```

Oy	81		-KOKPGDAAVKGEEIVAHNLIGVHDI	111
Dd	164	PLIDIKMSQVEQNPFPCDDPKDALDELMLAKLKEBD-SIGAKYIVASGVAGICEEV	222	
Oy	112	VS-MEYDLAVKLGDJHTTHV-ISDIODFVALSLSEISDEGNTTMTSFEEVROPANVNH	169	
Dd	223	FDRIDADIALHMSINMVKGEIGDGFDVVALNGSQNRDE--ITKOGFO-----SNHA	273	
Oy	170	GGL 172		
Dd	274	GGI 276		
RESULT 47 US-09-733-300-3 Sequence 3, Application US/09733300 Patent No. US20020061570A1 GENERAL INFORMATION: APPLICANT: Anderson, P.C. Chomet, P.S. Griffor, M.C. Kriz, A.L. TITLE OF INVENTION: ANTHRANILATE SYNTHASE GENE AND METHOD OF USE THEREOF FOR CONFERRING TRYPTOPHAN OVERPRODUCTION NUMBER OF SEQUENCES: 16 CORRESPONDENCE ADDRESS: ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A. STREET: P.O. Box 2938 CITY: Minneapolis STATE: MN COUNTRY: USA ZIP: 55402 COMPUTER READABLE FORM: MEDIUM TYPE: Diskette COMPUTER: IBM Compatible OPERATING SYSTEM: DOS SOFTWARE: FastSeq Version 2.0 CURRENT APPLICATION DATA: APPLICATION NUMBER: US/09/733,300 FILING DATE: 08-Dec-2000 PRIOR APPLICATION DATA: APPLICATION NUMBER: 09/312,721 FILING DATE: 2000-09-11 ATTORNEY/AGENT INFORMATION: NAME: Woessner, Warren D. REGISTRATION NUMBER: 30,440 REFERENCE/DOCKET NUMBER: 950.026US2 TELECOMMUNICATION INFORMATION: TELEPHONE: (612) 373-6903 TELEFAX: (612) 339-3061 TELEX: <unknown> INFORMATION FOR SEQ ID NO: 3: SEQUENCE CHARACTERISTICS: LENGTH: 595 amino acids TYPE: amino acid STRANDEDNESS: single TOPOLOGY: linear MOLECULE TYPE: protein SEQUENCE DESCRIPTION: SEQ ID NO: 3: US-09-733-300-3 Query Match 6.8%; Score 73; DB 10; Length 595; Best Local Similarity 21.2% ; Pred. No. 35; Matches 55; Conservative 30; Mismatches 84; Indels 90; Gaps 11;				
Oy	23	DITEELINKADDAIAIEQSETIDPMKVVDHADKFERHYGIYDFKGLAMRNIERGUK	82	
Dd	154	ETMHDEFEDPELMPIRKISEKMNDPOLYODLPDAF--CGGWGFSSYDVIRVRVERKKLP	211	
Oy	83	OMKROGDAVNGEGICVAHLLIGVHDII-----SMEDLAY----	120	

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12996
LENGTH: 6281
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match 6.9%; Score 74; DB 10; Length 6281;
Best Local Similarity 18.3%; Pred. No. 6.5e+02;
Matches 47; Conservative 44; Mismatches 80; Indels 86; Gaps 11;

QY 23 DKITEINKA--IDDAIAIEOSEITIDPMKVPDHDKFRHVGIVDFKGL-----71
DB 5169 NKASAINRATKLDNMGAVOO--YID-----EQHGVISSTNYINADNLKAN 5215
QY 72 -----AMRNEARGLKQMKRGDANVKGEGIV---KAHLIGVHD 109
DB 5216 YDNAIANAHLEDKVGNAIAKAEQKQNTIIDQNLNGQNLANAKDANAFVNSLN 5275
QY 110 DIVSEYDLAYKLGDPHTTHVISDIOFVALSLEISD-----148
DB 5276 GLNQOQDOLAHKAIN--NADTVSDVT-D-IVNQIDLDNAMEDTLKHLVDNEIIPNAEQYVA 5331
QY 149 ---EGNITMTSEVROFANV-----VNHIGGL--SLDPFGVLSVDLTATFQDPTV 194
DB 5332 YONADNNAKTNDDAKRLATILNSDNTNVNDINGAICAVNDAIHNLNGDQRLQDARKDA 5391
QY 195 RKEMTKVLAPARKLE 211
DB 5392 IQSINQALANKLK-EIE 5407

RESULT 41
US-09-996-634-135
Sequence 135, Application US/09996634
Patent No. US20020172684A1
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 61260
CURRENT APPLICATION NUMBER: US/09/996,634
PRIOR APPLICATION NUMBER: 09/447,135
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08/990,823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 135
LENGTH: 445
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-996-634-135

Query Match 6.9%; Score 73.5; DB 9; Length 445;
Best Local Similarity 22.6%; Pred. No. 22;
Matches 47; Conservative 40; Mismatches 66; Indels 55; Gaps 12;

QY 16 SADPHYDKITEINKAIDDAI-----AAIEO-SETIDPMKVPDHDKFRHVG 63
DB 126 TAVPEWDEVEKKEGLTRLADLSPAGELOQPLGAIAINQADYLD-----GNQDSL--HNA 178
QY 64 IYDFKGLAMRNEARG-----LKQMKRGDANVKGEGIVK-AHLLIGVHDDIYSMEYD 117
DB 179 LREL-AQVAGRLGDSRGDIFGTVMQVLDLSSDEQIVQAGHVASVQVLADSSAN 237
QY 118 LAAYKLGDPHTTHVISDIOFVALSLEISDEGNITM--TSFEVROFANV-----167
DB 238 LDQITGLTN---QALSDIRGLF-----RNNSTLIETVQNLDPQTLSDQSENIEQ 286
QY 168 --HIGG-----LSLDPFGVLSVDLT 187
DB 287 VLHVAGPGITNFEYNIYDPAQGTNLGLLS 314

RESULT 42
US-09-815-242-13637
Sequence 13637, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13637
LENGTH: 666
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13637

Query Match 6.9%; Score 73.5; DB 10; Length 666;
Best Local Similarity 21.3%; Pred. No. 37;
Matches 26; Conservative 23; Mismatches 36; Indels 37; Gaps 5;

QY 109 DDI---VSMEDYDLAYKLGDPHTTHVISDIOFVVA-----LSLEISDEGNITM 154
DB 422 DDOYKVSDDYDMVS-----TIKETEKPAYLVSVAMTEQAOIOLVELVAKDPNLEY 475
QY 155 TSFEVROFANVNHIGLSLDPFGVLSVDL-----TALQDPTVRKEMTKVLAPAK 207
DB 476 GDFEIOELNIVKR-----YGIIOELRLALKKNYLYKNDKREIVPVLEQILT 525

```
; SEQ ID NO 41
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-41
```

```
Query Match          6.9%; Score 74; DB 9; Length 928;
Best Local Similarity 24.0%; Pred. No. 51;
Matches 30; Conservative 20; Mismatches 31; Indels 44; Gaps 7;
```

```
QY 18 DPHVDKITEETINKAIDDA-----IAAIEOSETIDPKVP---DHADKFERH-- 61
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 EYVHGDFAFAQAVLDCILHRSABEGTLETARWLAVLEQS---DPLIVERLTGTTAAVERHIQ 311
QY 62 ----VGIVDFKG-----ELAMRNIEA-----RGLKOMKROGDANVKGEGIVKA 101
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 312 ELAATGILDEDEGTGCPAIRAALADLPAGERTELHRRAAEQULHRDG-----ADEDTVAR 366
QY 102 HLLIG 106
   |||:|
Db 367 HLLVG 371
```

```
RESULT 38
US-09-861-289-41
; Sequence 41, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-41
```

```
Query Match          6.9%; Score 74; DB 10; Length 928;
Best Local Similarity 24.0%; Pred. No. 51;
Matches 30; Conservative 20; Mismatches 31; Indels 44; Gaps 7;
```

```
QY 18 DPHVDKITEETINKAIDDA-----IAAIEOSETIDPKVP---DHADKFERH-- 61
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 EYVHGDFAFAQAVLDCILHRSABEGTLETARWLAVLEQS---DPLIVERLTGTTAAVERHIQ 311
QY 62 ----VGIVDFKG-----ELAMRNIEA-----RGLKOMKROGDANVKGEGIVKA 101
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 312 ELAATGILDEDEGTGCPAIRAALADLPAGERTELHRRAAEQULHRDG-----ADEDTVAR 366
QY 102 HLLIG 106
   |||:|
Db 367 HLLVG 371
```

```
RESULT 39
US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
```

```
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5834
; LENGTH: 2437
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5834
```

```
Query Match          6.9%; Score 74; DB 10; Length 2437;
Best Local Similarity 18.3%; Pred. No. 1.8e+02;
Matches 47; Conservative 44; Mismatches 80; Indels 86; Gaps 11;
```

```
QY 23 DKITEEINKA--IDDAIAAIEOSETIDPKVPDHDADKFERHVGIVDFKGL----- 71
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1327 NKASRAINRATRLDNAMGAVOO--YID-----EØHLGVISSTNYINADNLKAN 1373
QY 72 -----AMRNIEARGLKOMKROGDANVKGEGIV---KAHLIGVHD 109
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1374 YDMATNAAHELDKVOGNAIAKAEØLKNITIDQNALNGQONLANAKDKANAFNSLN 1433
QY 110 DIVSMETDLAYKLGDLHPTTHVYISDIQDFVALSLEISD----- 148
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1434 GLNQOQODLAHKAIN--NADRVSDVTD--IVNQØIDLNDAMETLKLVDNEIPNAEQFVN 1489
QY 149 ----ECNITMTSFEVRQGFNV-----VNHIGL--SILDPICGVLSDVLTATFODIV 194
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1490 YONADNMARTNEDDAKRLANTLNSDNTNVNDINGAIOAVNDAIHNLNGDØRLØDAKDKA 1549
QY 195 RREMTFVLAPAFKRELE 211
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1550 IOSINØALANKLK-EIE 1565
```

```
RESULT 40
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
```

```
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent ver. 3.0
; SEQ ID NO 3564
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3564
```

```
Query Match          7.0%; Score 74.5; DB 9; Length 285;
Best Local Similarity 26.4%; Pred. No. 9.6;
Matches 29; Conservative 22; Mismatches 38; Indels 21; Gaps 6;
```

```
OY 86 ROGDAVNGE-EGIVKAHLIG-VHDD-----IVSMEDYLAKLGLDHPPTHVISDIODF 138
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 126 RKLQDINSELGVISAGDKLGLDDBWTINIVGKVI--QIDARSTAFAD----- 179
OY 139 VVALSLEISDEGNITMTSEFVRQFANVNNHIGLSILDPFISVLDVLT 188
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 180 ----ATELINQANSEIVFPHVHMLNSKIVMSGFS---DYSLAGDAITS 221
```

RESULT 35

```
US-10-260-877-128
; Sequence 128, Application US/10260877
; Publication No. US20030021813a1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: 'ESSENTIAL GENES'
; FILE REFERENCE: 6565.US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128
; LENGTH: 357
; TYPE: PRT
; ORGANISM: H. influenzae
US-10-260-877-128
```

```
Query Match          7.0%; Score 74.5; DB 9; Length 357;
Best Local Similarity 18.6%; Pred. No. 13;
Matches 37; Conservative 34; Mismatches 55; Indels 73; Gaps 9;
```

```
OY 52 PHAD-KFERHGYIDFKG-----ELAMR-----NITARG----- 80
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 104 PGHADFTYQKXGIRYRGGRSSARETAMRVAGAIAKKYLREHFGIEVRGFSIQIGNI 163
OY 81 -----LKMROGDANVKGEGIVKAHLIG 106
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 164 KIAPOKVGQIDMEKVNNSPFPCDESAYVEKFEELIRELKESD-SIGAKLYIAENVPYG 222
OY 107 VHDDIVS-MEYDLAYKLGDLHPTTHVISDIOQ-FVVALSLEISDEGNITMTSEFVRQFAN 164
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 223 LGEYPEDRLADADLAHALMGINAVKV--EIGDGFANVEGRGSEHREDEMPNGFESNHAGG 280
OY 165 VVNHIGLSILDPFISVLS 183
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 281 I---LGGISSGQPIATIA 296
```

RESULT 36

```
US-09-815-242-10981
; Sequence 10981, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/1191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10981
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10981
```

```
Query Match          7.0%; Score 74.5; DB 10; Length 357;
Best Local Similarity 18.6%; Pred. No. 13;
Matches 37; Conservative 34; Mismatches 55; Indels 73; Gaps 9;
```

```
OY 52 PHAD-KFERHGYIDFKG-----ELAMR-----NITARG----- 80
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 104 PGHADFTYQKXGIRYRGGRSSARETAMRVAGAIAKKYLREHFGIEVRGFSIQIGNI 163
OY 81 -----LKMROGDANVKGEGIVKAHLIG 106
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 164 KIAPOKVGQIDMEKVNNSPFPCDESAYVEKFEELIRELKESD-SIGAKLYIAENVPYG 222
OY 107 VHDDIVS-MEYDLAYKLGDLHPTTHVISDIOQ-FVVALSLEISDEGNITMTSEFVRQFAN 164
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 223 LGEYPEDRLADADLAHALMGINAVKV--EIGDGFANVEGRGSEHREDEMPNGFESNHAGG 280
OY 165 VVNHIGLSILDPFISVLS 183
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 281 I---LGGISSGQPIATIA 296
```

RESULT 37

```
US-09-860-846-41
; Sequence 41, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
```


Oy 198 MTKVLAPAFKRELEK 212
 Db 165 ----LTPSDTAOLNK 175

RESULT 32

US-09-815-242-12351
 ; Sequence 12351, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes In
 ; FILE REFERENCE: ELTRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO: 12351
 ; LENGTH: 572
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-12351

Query Match 7.0%; Score 75; DB 10; Length 572;
 Best Local Similarity 21.0%; Pred. No. 22;
 Matches 41; Conservative 40; Mismatches 72; Indels 42; Gaps 9;

Oy 32 AIDDAIAIEQSETIDPMKVPDHPADKFERHVGIVDPFKGLAMRN--IEARGLKOMKROGD 89
 Db 9 AASDGAIAIKAKALLVEPDLT-----FDKNEKVTDEGEVAKFNIEISKVELITKIRNN 62
 Oy 90 ANVK---GEEGIVAKAHLI-----GVHDDIVSMEDYLAAYKLGDLHPPTHVHISDIODF 138
 Db 63 ABEVQAGADRAAIFDAHLVLDPELLOPIODKIKNNANAAATLTDV--TQGFVT----- 115
 Oy 139 VVALSLEISDEGNITMTSEVROFA-NVYNHITGSLIDPIGVLSDVLTALFODIVRKE 197
 Db 116 ----IPESMDNEYMKERRADIRVSKRVLSHILGVELPMP--SMIDESVVIIGND----- 164
 Oy 198 MTKVLAPAFKRELEK 212
 Db 165 ----LTPSDTAOLNK 175

RESULT 33
 US-09-728-137-2
 ; Sequence 2, Application US/09728137
 ; Patent No. US20010029031A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walter F. Boron

APPLICANT: Antoine Michel Alain Brill
 APPLICANT: Nassirah Khandoudi
 APPLICANT: Xavier Martin
 APPLICANT: Steven Charles Juge
 APPLICANT: Christopher John Rawlings
 APPLICANT: Trudy Rachel Doe
 ; TITLE OF INVENTION: NOVEL COMPOUNDS
 ; FILE REFERENCE: GH-30409-D1
 ; CURRENT APPLICATION NUMBER: US/09/728,137
 ; CURRENT FILING DATE: 2000-12-01
 ; PRIOR APPLICATION NUMBER: 09/115,954
 ; PRIOR FILING DATE: 1998-07-15
 ; PRIOR APPLICATION NUMBER: EP 97401714.7
 ; PRIOR FILING DATE: 1997-07-16
 ; PRIOR APPLICATION NUMBER: EP 97401713.9
 ; PRIOR FILING DATE: 1997-07-16
 ; PRIOR APPLICATION NUMBER: EP 98400272.5
 ; PRIOR FILING DATE: 1998-02-09
 ; PRIOR APPLICATION NUMBER: EP 98400454.9
 ; PRIOR FILING DATE: 1998-02-26
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO: 2
 ; LENGTH: 1044
 ; TYPE: PRT
 ; ORGANISM: HOMO SAPIENS
 US-09-728-137-2

Query Match 7.0%; Score 75; DB 10; Length 1044;
 Best Local Similarity 20.1%; Pred. No. 48;
 Matches 45; Conservative 38; Mismatches 79; Indels 62; Gaps 10;

Oy 35 DAIAIEQSETIDPMKVPDHPADKFERH---VGIVDFKGLAMRNIEARGLKOMKROGD 89
 Db 19 DEEAIVDQSGCTIILNINHEKELECHRTLYGVGNRPLRGSRHNRHTGOKHNRGRGRK 78
 Oy 90 ANVKGEIGI-----VKANHLIGVHDDIVSMEDYLAAYKGLD-----HPTTH 130
 Db 79 GASQGEGLALAHPTSPQROVFICTEEDDEHVPHELPTLEIDICMKGEDEAKMETAR 138
 Oy 131 VI---SDIOD-----FVALSLEISDEGNITMTSEVR-----QFANVYNN 168
 Db 139 WKFEEDVEDGGERMSKPVATLSLH-----SLFELRCLNGTIVLLDMHANSIEE 189
 Oy 169 IGLSLIDPIFGVLSVDLTAIFODIVRKEMTKVLAPAFKRELEK 212
 Db 190 ISDL-ILD-----QOELSSDLNDSMKRVKREALLKKNHONK 226

RESULT 34
 US-09-738-626-3564
 ; Sequence 3564, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHITAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988

```
FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13511
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13511

Query Match
Best Local Similarity 7.1%; Score 76; DB 10; Length 567;
Matches 58; Conservative 38; Mismatches 81; Indels 54; Gaps 14;

OY 1 MMKFLIAAFAVAVSADFIHYDKIT--EEINKAIDAIAIEQSETIDPMKVPDHADK 57
DB 329 VVKYLL--ANGF-----LHGDRITCTGKTVAEMLADPADLTPGQKVIEMLENPKRADG 379
OY 58 FERHGVDFKGLAMRNIEAR--GIKQKROGDANV--KGEIGYKAHLILVHD--DIYS 113
DB 380 -----PLILGNLAPDCAVAVSGVKVRHVPKAVPDSDDAIOAVLTDEIVGDVVV 434
OY 114 MEYDLAYKIGDLHPTTHVISDIQDEVALSLEISDEGN-----ITMTSEYEROFANVYN 167
DB 435 VRF-VGPKGGPGMP-----MLSLSMIVGGQGDKVALITDGRFSGTGYLVYG 483
OY 168 HI-----GGLSLDPIRGVSLDTLAIPTQDTRKEMTKVLAPARELEK 212
DB 484 HIAPADGG-----PLAYLRTGDIIVTVDDPT--KEISMAVS---EEELEK 524

RESULT 30
US-09-905-983-58
; Sequence 58, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadoz, Elia
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-905-983-58

Query Match
Best Local Similarity 7.1%; Score 75.5; DB 10; Length 905;
Matches 47; Conservative 48; Mismatches 85; Indels 65; Gaps 9;

OY 18 DEIH--YDKIE-----EINKAIDAIA-----AIDQSETIDPMKVPDHAD 56
DB 360 DPLNIAIDKMTKTRDLRQLRKAVMDHISDFLETNVPLVLIVFAKSGNEKVEYKVAQ 419
```

```
OY 57 KEERHVG-IVDFKGLAMRNIEARGLKQK-----RQ 87
DB 420 VFERHANKLVEANLACSI--NNBEGVKLVMAATQIDSLCPVOYNALTLARPSKVAQ 479
OY 88 GDANKKEEGIVKKAHLIGVHDDIVSMEDLAYKIGDLHPTTHVISDIQDEVALSLEIS 147
DB 480 DNMDFKQDQEKQVRVLTFRAVDITSDDFLSVS-----ENHILEDVKKCVIA--LQBG 531
OY 148 DEGNITMTSEFYEROFANVYNIHIGLSIIDPIRGVSLD-VLTAIFQDTYKREMTKVLAPAF 206
DB 532 DVDITDRTAGAIRGRAARVHIHINEMENYEAQVTEKYLEA-----TKLSEFVMPRF 585
OY 207 KRELE 211
DB 586 AEQVE 590

RESULT 31
US-09-815-242-5526
; Sequence 5526, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5526
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5526

Query Match
Best Local Similarity 7.0%; Score 75; DB 10; Length 572;
Matches 41; Conservative 40; Mismatches 72; Indels 42; Gaps 9;

OY 32 AIDDAIAIEQSETIDPMKVPDHADKFEHNVGIVDFKGLARN--IEARGLKQKROGD 89
DB 9 AASDGVAILAKAVLLVPEPDLT-----FDKNEKRVTDGEVAFENSAIEASKVELTKIRRN 62
OY 90 ANVK---GEEGIVKKAHLI-----GVHDDIVSMEDLAYKIGDLHPTTHVISDIQDE 138
DB 63 AEYQUGADKAALFADHLLVLDPDELIOPTQDKIKMENMAAATALDV--TTPFVY----- 115
OY 139 VVALSLEISDEGNITMTSEFYERQFA--NVYNHIGLSIIDPIRGVSLDTLAIPTQDTRKE 197
DB 116 ----IFESMDNMYMKERADIDVSKRVLSHLIGVELPMP--SMIDESVVIYGND----- 164
```

```
Db 50 MKVINKSILSRAAKYKGLDGEVFGPTAVAFSNDVVAAPAKIIDEFKAD---AKALEI 106
Oy 147 SD---EGNTTMTSFYKROANVNNHIGLSILDPITFGVLSVDLTAIFODTVRKEMTKVLA 203
Db 107 KCGVIEGKVSVE-OTTLAKLAPNREGLLS-----MLSLVLAQAPVRN-----VA 149
Oy 204 PAFKRELEKN 213
Db 150 YAVKAVAEKN 159

RESULT 27
US-09-815-242-10614
; Sequence 10614, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10614
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10614

Query Match 7.2%; Score 76.5; DB 10; Length 166;
Best Local Similarity 26.8%; Pred. No. 3;
Matches 51; Conservative 32; Mismatches 62; Indels 45; Gaps 13;

Oy 38 AAIEOSETIDPMKVPDHAKFE--RHVGIVDFKGLAMRNIEARKGRGODAVKGE 95
Db 4 AAIARETIL---VQAARKFESASAVIVDRG---LVEEVTNLRKQLR--DAV--E 52
Oy 96 EGIYVAHLL-----IGVH--DDIVSMEDLAYKIGLDLHPTTHVISDI--QDFVALSLEI 146
Db 53 MKVINKSILSRAAKYKGLDGEVFGPTAVAFSNDVVAAPAKIIDEFKAD--AKALEI 109
Oy 147 SD---EGNTTMTSFYKROANVNNHIGLSILDPITFGVLSVDLTAIFODTVRKEMTKVLA 203
Db 110 KCGVIEGKVSVE-OTTLAKLAPNREGLLS-----MLSLVLAQAPVRN-----VA 152
Oy 204 PAFKRELEKN 213
Db 153 YAVKAVAEKN 162
```

```
RESULT 28
US-09-738-626-3637
; Sequence 3637, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAMA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3637
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3637

Query Match 7.1%; Score 76; DB 9; Length 308;
Best Local Similarity 23.0%; Pred. No. 7;
Matches 56; Conservative 28; Mismatches 80; Indels 80; Gaps 13;

Oy 6 LIAAVFVAVSAD---PIHYDKITEINKAIDDAIAIEOSETIDPMKVPDHAKFEERH 61
Db 97 LVVGLEPIIAISVDYRLAPH-----PPRAIIDDAAFAVVS-----AVLDGVSGLSID 142
Oy 62 VGIYDFKGLAMRNIEARKLQMKRGGDANVKGEGIVAKHLLIGVHDIV----- 112
Db 143 TFRVAIGDSAGGINAAVAAOQLRER---AVGSTVLAHQVLIFFVTSTSTPSLYLT 198
Oy 113 -----SME-----YD-----LAKKIGLDLHPTTHVISD--IQDFVYA 141
Db 199 FSKDCYLTKDAMERYIEQYADGHDRTDPRLSPLLASDLSDLPTTVVGECDVLAHEVRA 258
Oy 142 LSLAISDEGN-ITMTSF--EVROFANVNNHIGLSILDPITFGVLSVDLTAIFODTVRKEM 198
Db 259 YGQALLAENGNSVTMTFFKQIHAFIN---LGSIS-----SDARAA--RLIRARL 303
Oy 199 TKVL 202
Db 304 EAAL 307
```

```
RESULT 29
US-09-815-242-13511
; Sequence 13511, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
```

```

:
:       TOPOLOGY: linear
:       IMMEDIATE SOURCE:
:       LIBRARY: GENE BANK
:       CLONE: g133985
:       SEQUENCE DESCRIPTION: SEQ ID NO: 3 :
US-09-828-423-3

Query Match              7.3%; Score 77.5; DB 10; Length 946;
Best Local Similarity 19.8%; Pred. No. 24;
Matches 57; Conservative 46; Mismatches 84; Indels 101; Gaps 14;

QY 19 PIHYDKTEIEINKAID--DAIAIEQSEITIDPMKYDPDHAKDFERRHVGIVDFKGEI-AMRN 75
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 301 PDLNIPKPNILFEVLDVSGSMGVKMKQTVKAMKTIIDDLRAEDHPSVIDFNONIRTRMN 360
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 76 -----IEARGKQMKROGDANKGEGIVKAHLILGVHDDIVSME---YDLAK 121
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 361 DLFOLOKHLQIAKRYIEIKIOPSGGTNI--NEALLRAIFILNEANNLGLDPNSVSLIL 418
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 122 LGDLHPTT-----HVISIDIDFVVALSL-----EISDE-----G 150
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 419 VSDGPTVGEIQLSKIQKVKENIDNISLFSLGMGFDVDYDFLKRISNNHGIAORIG 478
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 151 NITMTSEFEVROPANV-----NHIIGLSIL-----DPT- 178
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 479 N-QDTSSQLKRFYNOVSPELRNQGPNYHPSVTQYQNNFNHYGSGSEIYVAGKRPDAK 537
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 179 FEVLSDVLTAI-----FQDLYRKEKTKYLAFAKREL 210
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 538 LDQIESVIYATSAANTQVLETLAQDDLDODFLSKD--KHADDPFRKL 583
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 25
US-09-815-242-11000
: Sequence 11000, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11000
: LENGTH: 722
: TYPE: prt
: ORGANISM: Haemophilus influenzae
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(722)
```

```

:       OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-11000

Query Match              7.2%; Score 77; DB 10; Length 722;
Best Local Similarity 21.9%; Pred. No. 19;
Matches 43; Conservative 41; Mismatches 86; Indels 26; Gaps 9;

QY 7 IAAVFVAVSADPIHYDKTEIEINKAIDDAIAIE--QSEITDPMKYDPDH-----ADKFER 60
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 526 IASIVAVSINDEPIAFALTDITLK---NDSLHAIORLOQNDIVYIMSGDQSVVDYIAK 582
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 HVGIVDFPKGLAMRNIEARGKQMKROGDANKGEGIVKAHLILGVHDDI-VSMEYDLA 119
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 583 QLGIKKAFQGLTPRD-KAQIQIKLDGLGHIVAMVGDINDAPALASAVSEFAMKSSSDIA 641
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 120 YKLGDLHPTTHYISDIQ--FVALSLSEISDEGNTMTSFEVROPANVNHIG---GLS 173
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 642 EGTASATIMQHSVNOLVVALFIARATLK---NIKQMLE---FALIYNILGIPLAAMFG 692
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 174 ILDPIFEVLSDVLTAI 189
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 693 FLSPILAGAAALSSI 708
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 26
US-09-815-242-4914
: Sequence 4914, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4914
: LENGTH: 161
: TYPE: prt
: ORGANISM: Enterococcus faecalis
US-09-815-242-4914

Query Match              7.2%; Score 76.5; DB 10; Length 161;
Best Local Similarity 26.8%; Pred. No. 2.9;
Matches 51; Conservative 32; Mismatches 62; Indels 45; Gaps 13;

QY 38 AAIEQSEITIDPMKYDPDHAKFE--RHVGIVDFKGLAMRNIEARGKQMKROGDANKGE 95
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1 AATAKKREL---VQAARKEFESASVIVDYRG--LVVEEYTNIRKQDLR--DAGV--E 49
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 96 EGIVKAHL-----IGVH--DDIVSMEYDLAVKLGDLHPTTHVISDI-QDFVVALSLERI 146
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```



```
Db 504 DTLHERVIGQKDVANSISKAVRRARAGLK-----DP-----:KPRIGSFILGPTG 548
Qy 73 MRNIE-ARGL-KOMKROGDANVKE-EGIVKAHL---LIGV-----HDDIVSMEDLAY 120
Db 549 VGKTELAALAESMGDDAMIRVMSFEMEKHAVSRVAGAPGVGHDDGQLTEKRYR 608
Qy 121 KLGLDHPHTHVISD----IQDFVALSLEISDEGNITMTSEFVRQFANVY---NHIGGL 172
Db 609 K-----PYSVILFDEIEKAHPDVFENILLQVLDGDHLTDTKGRVDFRMTIIMTSVGAQ 663
Qy 173 SLIDPIFGVLSVLTAFODTVRKEMTKVLAFAFKRE 209
Db 664 ELQDQRFAGFGSSSDGQDYETIRKTMLEKNSFRPE 700

RESULT 20
US-09-815-242-12789
; Sequence 12789, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12789
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12789

Query Match 7.4%; Score 78.5; DB 10; Length 818;
Best Local Similarity 22.6%; Pred. No. 16;
Matches 49; Conservative 37; Mismatches 86; Indels 45; Gaps 11;

Qy 18 DPHYDKTTEE-----INKAIDATAIEQSETIDPMKVPDHDKFERHVGIVDEKGEIA 72
Db 504 DTLHERVIGQKDVANSISKAVRRARAGLK-----DP-----:KPRIGSFILGPTG 548
Qy 73 MRNIE-ARGL-KOMKROGDANVKE-EGIVKAHL---LIGV-----HDDIVSMEDLAY 120
Db 549 VGKTELAALAESMGDDAMIRVMSFEMEKHAVSRVAGAPGVGHDDGQLTEKRYR 608
Qy 121 KLGLDHPHTHVISD----IQDFVALSLEISDEGNITMTSEFVRQFANVY---NHIGGL 172
Db 609 K-----PYSVILFDEIEKAHPDVFENILLQVLDGDHLTDTKGRVDFRMTIIMTSVGAQ 663
Qy 173 SLIDPIFGVLSVLTAFODTVRKEMTKVLAFAFKRE 209
```

```
Db 664 ELQDQRFAGFGSSSDGQDYETIRKTMLEKNSFRPE 700

RESULT 21
US-09-738-626-4574
; Sequence 4574, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4574
; LENGTH: 1209
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4574

Query Match 7.4%; Score 78.5; DB 9; Length 1209;
Best Local Similarity 22.3%; Pred. No. 27;
Matches 39; Conservative 33; Mismatches 68; Indels 35; Gaps 6;

Qy 13 VAVSADPHYDKITEINKAIDATAIEQSETIDPMKVPDHDKFERHVGIVDEKGEIA 72
Db 45 LVVATATGREAEDLTALAKAMGDKVAMLPSEWTLPHERLSPGVDIYGRACVLAHNLGDLK 104
Qy 73 MRNIEARGLKO---MKROGDANVKEEG-----IYKAHLILGVHDDIVSMED 117
Db 105 VVVAARAFPCQVYLKDAEGRAPLTKEGAEPDFSAITTELVEFRAY---KHVDVAKKGE 160
Qy 118 LAYKLG--DLHPTT-----HVISDQDFVALSLEISDEGNITMTSFEV 159
Db 161 FATRGGIIDFPPTLTIDYPRVREFWGDVSDIRQFSVAQRTIPE---ITIKSIEI 212

RESULT 22
US-09-815-242-12956
; Sequence 12956, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
```

APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5843
LENGTH: 550
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5843

Query Match 7.4%; Score 79; DB 9; Length 550;
Best Local Similarity 27.0%; Pred. No. 8.5;
Matches 51; Conservative 29; Mismatches 61; Indels 48; Gaps 11;

QY 5 LLIANAAYVAVSADPIHKDITEEINKAIDAIAIEOSETIDPMKVPDHAUKFERHVG 63
DB 88 LIDAAVAPVRAVDA-----AITDLSTKIDGASLSAEQAOV-----ATDFDALG 132
QY 64 IYFKGELAMRNTEAR-----GLKQMKRQ---GDANKGEGIVKAHLIGVHDIY 112
DB 133 AAE---ELGWEIDARIEITVAGLAEVDRSPPELSCGRRRALALLLEPHDALI 189
QY 113 SMEYDLAYKIGDLPTTHVVISDIOFVALSLDISD-EGNITMTSFEVROFANVNHIG 171
DB 190 FDE-----PINHLDDTAVDFLIS---DISRFGVLLASHD-RFFLDV---CTE 232
QY 172 LSLIDPIFG 180
DB 233 LIDLDPALG 241

RESULT 18
US-09-815-242-5586
Sequence 3586, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5586
LENGTH: 810
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5586

Query Match 7.4%; Score 78.5; DB 10; Length 810;
Best Local Similarity 22.6%; Pred. No. 16;
Matches 49; Conservative 37; Mismatches 86; Indels 45; Gaps 11;

QY 18 DPHYDKITEE-----INKAIDAIAIEOSETIDPMKVPDHAUKFERHVGIVDFKGEIA 72
DB 504 DTLHERVIGCKDAVNSISRAVRRARAGLK-----DKPISFIFLGPFG 548
QY 73 MRNIE-ARGL-KMKRQGANVKG-EGIVKAHL---LIGV-----HDDIVSMEVDLAY 120
DB 549 VGRTELARALAESMGDDMIKVDMSFEMEKAVSRVLGAPPGYVGHDDGGLTEKVR 608
QY 121 KLGDLPHTTHVISD-----IODEFVALSLDISDEGNITMTSFEVROFANV---NHIGGL 172
DB 609 K-----PYSVILFDELEKAPVDVFNLLQYDDGHLTDKGRVDFRNTIIMTSNVGMA 663
QY 173 SLIDPIFGVLSVDVLTAFODTVRKEMTKVLAPEKRE 209
DB 664 ELQDORFAGCGSSDQDVEYETIRKTMKELKNSFRPE 700

RESULT 19
US-09-815-242-12424
Sequence 12424, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12424
LENGTH: 818
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12424

Query Match 7.4%; Score 78.5; DB 10; Length 818;
Best Local Similarity 22.6%; Pred. No. 16;
Matches 49; Conservative 37; Mismatches 86; Indels 45; Gaps 11;

QY 18 DPHYDKITEE-----INKAIDAIAIAIEOSETIDPMKVPDHAUKFERHVGIVDFKGEIA 72

```
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5468
; LENGTH: 831
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5468
```

Query Match	7.5%;	Score 80;	DB 9;	Length 831;
Best Local Similarity	24.4%;	Pred. No. 12;		
Matches 43;	Conservative 31;	Mismatches 70;	Indels 32;	Gaps 8;

```

Qy 29 INKRIIDAIIAIOSQETIP---MVPHPADKFEF-----HGVIDFK---G
Db 97 INLVNDIHSHSVTPPELDSVLGVIPD-ADSFDMDEFDLLRNKAKMHHGVAAIEPDGQ
Qy 70 ELAMRNTEALGTLQMRGDDAVNKGEEGLVKAHLLIGHD--DIVSMEDLAKKLDLHR
Db 156 ELVYEVKPKA-----NSASRQDISEGVDLSTGAADLEADVEDEV-TAFPLEYLNK
Qy 128 TTHVYISDIQDFVALSLSEISDGRNTMTMSFEYRQFANVANHIGSLIDLPFGVLS 183
Db 207 TSEVDIDDEGCVDPGLSEGVNNAAG-ITDDLLIRELIDVDPSSAAELINDINIVFT 261

```

RESULT 15
US-09-815-242-10901
; Sequence 10901, Application US/09815242
; Patent No. US20020061569A1

1 APPLICANT: Haselbeck, Robert
 2 APPLICANT: Ohlsen, Karl L.
 3 APPLICANT: Zysekind, Judith W.
 4 APPLICANT: Wall, Daniel
 5 APPLICANT: Trawick, John D.
 6 APPLICANT: Grant, J.
 7 APPLICANT: Yamamoto, Robert T.
 8 APPLICANT: Xu, H. Howard
 9 TITLE OF INVENTION: Identification of Essential Genes in
 10 TITLE OF INVENTION: Prokaryotes
 11 FILE REFERENCE: ELITRA.011A
 12 CURRENT APPLICATION NUMBER: US/09/815,242

```

1 PRIOR APPLICATION NUMBER: 60/191,078
2
3 PRIOR FILING DATE: 2000-03-21
4
5 PRIOR APPLICATION NUMBER: 60/206,848
6
7 PRIOR FILING DATE: 2000-05-23
8
9 PRIOR APPLICATION NUMBER: 60/207,727
10
11 PRIOR FILING DATE: 2000-05-26
12
13 PRIOR APPLICATION NUMBER: 60/242,578
14
15 PRIOR FILING DATE: 2000-10-23
16
17 PRIOR APPLICATION NUMBER: 60/253,625
18
19 PRIOR FILING DATE: 2000-11-27
20
21 PRIOR APPLICATION NUMBER: 60/257,931
22
23 PRIOR FILING DATE: 2000-12-22
24
25 PRIOR APPLICATION NUMBER: 60/269,308
26
27 PRIOR FILING DATE: 2001-02-16
28
29 NUMBER OF SEQ ID NOS: 14110
30
31 SOFTWARE: FASTSEQ For Windows Version 4.0
32
33 SEQ ID NO 10901

```

ORGANISM: Enterococcus faecalis
US-09-815-242-10901

Query Match	7.4%	Score 79.5;	DB 10;	Length 718;
Best Local	19.9%;	Pred. No. 11;		
Matches 45;	Conservative 36;	Mismatches 98;	Indels 47;	Gaps 8;

```

Qy      5  LLIAAVFAVVASDPRIHYDKTTEETINKAIDDAIAIQSEITIDPMKVPDHADKFEHHVGI  64
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      345 LAIGGLAFAMSGKDEVPDYVTNETKKAASQALQSAGIKYDSETKKIPD--DKLE-----  397

Qy      65 VDFKGLAMRIIEARGLKQMKR-----QGDAVWGKEEIVYKAHLLIGYHDD  110

```

```

Db      398  ---ECKVVKTPPEAKSSVKKKRSTVLIYSCTEKIEIEMADITNEYESAVELKKLGISED 45
QY      111  IVSMEYDLAYKLGDLHPTTHVIS-----DIODEFVALSLEISDEGNITMTSFEVRQ 161
Db      455  QITTRKEYS---DSVSTDNIIIMQKPAAGKVKDEKDKKVLTVSECPKA-VLPSSVAGYS 509
QY      162  FANVYNHIIGGSLIDPIFGVSLDVLTAIFQ--DTVRKEMTKVLAPA 205
Db      510  YTNAAVNALALQ-----GISISOITRVDQASDPTVEGVLTTODPA 548

```

```

RESULT 16
US-10-147-026-16
; Sequence 16, Application US/10147026
; Publication NO. US20030003538A1
; GENERAL INFORMATION:
; APPLICANT: Dietrich, Paul Shartzer
; APPLICANT: Huang, Chiao-Chain
; APPLICANT: Johnson, Carl D.
; APPLICANT: Sanganesswaran, Lakshim
; TITLE OF INVENTION: Neuropathic Pain Genes; Compositions
; TITLE OF INVENTION: Theoreof; Related Reagents
; FILE REFERENCE: ROCH-006
; CURRENT APPLICATION NUMBER: US/10/147,026
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/155,702
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/189,931
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 6304
; TYPE: PRT
; ORGANISM: human
US-10-147-026-16

```

Query Match	7.48;	Score	79.5;	DB	9;	Length	6304;
Best Local Similarity	20.98;	Pred. No.	2e+02;				
Matches	41;	Conservative	32;	Mismatches	70;	Indels	53;
						Gaps	8;

[illegible]

RESULT 17
US-09-738-626-5843
Sequence 5843, Application US/0973862626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIJO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASARO

OTHER INFORMATION: Xaa = Any Amino Acid
US-09-874-923-110

Query Match 7.8%; Score 83; DB 10; Length 845;

Best Local Similarity 23.3%; Pred. No. 6.3;

Matches 42; Conservative 34; Mismatches 70; Indels 34; Gaps 10;

QY 29 INKAIDDAIAIEOSEITIDPMKVPDADKFERHVGIVDFKGLAMNIEARGLKOKKROG 88
DB 2 VNTVDQVVELMDYDPQINMSYIAHVD-----HKSITLSDSLVCAGAGIKKEEAG 52
QY 89 DANV---KKEEGIVKAHLIGVHDDIVSMEDLAYK-LGDLHPHTTVISDIQDFVALSL 144
DB 53 DKRIMPTRADE-IARG---ITIKSTAIISMHYHPKEMIGDLO-----DDKRFDLINL-- 100
QY 145 EISDEGINMTSFVEVROFANVYVNHIGSLIDPIFGVLSVLAIFODVYRKEMTVLAP 204
DB 101 -IDSPGHVDFSS-EVTAALRYVD--GALVYVDCVEGVCVQETETVL-----RQALTERIRP 151

RESULT 12
US-09-841-132-399

Sequence 399, Application US/09841132

Patent No. US20020061848A1

GENERAL INFORMATION:

APPLICANT: Bhalla, Ajay

APPLICANT: Skelky, Yasir A.W.

APPLICANT: Probst, Peter

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

FILE REFERENCE: 210121.469C8

CURRENT FILING DATE: 2001-04-23

NUMBER OF SEQ ID NOS: 599

SOFTWARE: FastSeq for Windows Version 3.0/4.0

SEQ ID NO 399

LENGTH: 461

TYPE: PR

ORGANISM: Chlamydia pneumoniae

US-09-841-132-399

Query Match 7.7%; Score 82; DB 10; Length 461;

Best Local Similarity 20.4%; Pred. No. 3.5;

Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;

QY 23 DKITEE-----INKAIDDAIAIEOSEITIDPMKVPDADKFE-----RH 61
DB 219 NKFTKOGINILTKA---SISALEESONQVRIYVDQVEEDVYVAIGROFNASTIGLON 275
QY 62 VGIIV-DFKGLAMNIEARGLKOKKROG-----ANVKGEGIVKAHLIGVHDDIVS 113
DB 276 AGVIRDRDGVIPVDETMRNVPIVAIGDITGKWLAAHVASHGVAIAAKNSIHHN---V 332
QY 114 MEYDLAYKGGDLHPHTTVISDIQDFVALSLSEISDEGNT--TMTSEVRO----- 161
DB 333 MDYSAIPSVIETFR-----EIAVVGSLDEAEQONLPAKLTKEPKFAIGKAVALGAS 384
QY 162 --FANVYVNH-----IGLSLIDPIFGVLSVLAIFODVYRKEMTV 199
DB 385 DGFALIVSHETQOILGAVYVIGHASSLIGEMTL---AIRNELT 425

RESULT 13
US-09-738-626-5445

Sequence 5445, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5445
LENGTH: 785
TYPE: PR
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5445

Query Match 7.7%; Score 82; DB 9; Length 785;

Best Local Similarity 23.6%; Pred. No. 7.1;

Matches 63; Conservative 38; Mismatches 80; Indels 86; Gaps 17;

QY 7 IAAVAVVASADPI--HYKITE-EINKAIDDAIAIEOSEITIDPMKVPD--ADKFERHY 62
DB 58 IAVLSTNAADITAKNKVTSMTISKVHEIYA-----HNPDPHEISTIDTIT 107
QY 63 GIYDFK-GEIAMRN---IARGL--KOMKROGDANYKGEIGIVKAHL--LIGVHDDI--V 112
DB 108 NPLIDIEGQWMTSDYMIQDLRLDLYVMTGOGANLTAISFESHIEAFISVLDQIKGT 167
QY 113 SME-----YDLAYKGGDLH--PTTVISDIQD-----FVAL-----SLEISDEGN 151
DB 168 SLEIITICYLKDLKLEPHASPKYLLIDDEVODNSVEFEVFAIRFAKHNTSLYLVDSS 227
QY 152 ITMTSF-----EVROFANV-----VNHIGLSITL 175
DB 228 QTLIFERSANPKALNSLEASGVFGYRLTNTNRNOEIIIDFANIHSIDIEANOFAGIOLY 287
QY 176 DPLFGVLSVLA--IFODVYRKEMTV 201
DB 288 ANSF-----DAPVADSPFEKVEYELDMHHV 310

RESULT 14
US-09-738-626-5468

Sequence 5468, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO

APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

[illegible]

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5470
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5470

Query Match      8.1%, Score 86.5; DB 10; Length 660;
Best Local Similarity 22.2%; Pred. No.2.1;
Matches 36; Conservative 31; Mismatches 50; Indels 45; Gaps 5.

OY    1 MWKFLIAAVFVAASDPIDHYDKITTEENKAIDAIAIEGSETIDPMKV-----51
       |::| |::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB     354 IFSLMIALVSVAAAMGNKYEEIPDVIIGSKVKRA-----EOLFKNNLIKLKRISRSYSD 409
              :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY     52 -----PDHAKTER--HGIVDFKG-----ELARNRIEARGLNQW 84
              |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB     410 KYPEEIETPTPNTGERVERGDSDVDVISKGEKVKMPNVIGLPKEQALOKLSGLKDVG 469
              |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY     85 KRQGDAENVKGEGT-----YKMHLLICGHDDLVSMETYDLAYK 121
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB     470 KIEKYNNQAPRGGYIANOSVTANTEIATIHDSNIKIYESLGIR 511
               |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 8
US-09-815-242-12179
; Sequence 12179, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

Db 103 EXINSLNDLVKVEKSEVNGSPSSVYTRILKSQHAAMF-RHFOQIMFYNDITAAKQEK 161
Qy 71 ---LAMRNIEARG-----LKQMKRGDANKGEE-----GIVRAHL--LIGVHDDIVS 113
Db 162 CFTFLRQLEAVAGKEMSEEDVNDMLHOGKWEVFNESLLEINITYKAQISEIQRKHELVN 221
Qy 114 MEYDLAYKLGIDLPHTTVISDIQDFVALSLSEISDEG-----NITMTSEVROFANVANH 169
Db 222 LE-----NQIKDLRDLFIQISLVEEGESINNEMTVNSTKEYVNNTRK 267
Qy 170 GGLSI 174
Db 268 FGLAV 272

RESULT 4

US-09-898-570-30
; Sequence 30, Application US/09898570
; Patent No. US20020123612A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: EILERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; FILE REFERENCE: 15966-776CIP
; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/215,855
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/839,446
; PRIOR FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 30
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: POLYX
; OTHER INFORMATION: hnh0778p17_A1
US-09-898-570-30

Query Match 9.0%; Score 96; DB 10; Length 294;
Best Local Similarity 23.8%; Pred. No. 0.089;
Matches 44; Conservative 31; Mismatches 58; Indels 52; Gaps 10;
Qy 27 EEINKAIDAIAROSEID-PMKY-----PDHADKFERHVGIVDF-----KGE-70
Db 103 EXINSLNDLVKVEKSEVNGSPSSVYTRILKSQHAAMF-RHFOQIMFYNDITAAKQEK 161

Qy 71 ---LAMRNIEARG-----LKQMKRGDANKGEE-----GIVRAHL--LIGVHDDIVS 113
Db 162 CFTFLRQLEAVAGKEMSEEDVNDMLHOGKWEVFNESLLEINITYKAQISEIQRKHELVN 221
Qy 114 MEYDLAYKLGIDLPHTTVISDIQDFVALSLSEISDEG-----NITMTSEVROFANVANH 169
Db 222 LE-----NQIKDLRDLFIQISLVEEGESINNEMTVNSTKEYVNNTRK 267
Qy 170 GGLSI 174
Db 268 FGLAV 272

RESULT 5

US-09-898-570-32
; Sequence 32, Application US/09898570
; Patent No. US20020123612A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: EILERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; FILE REFERENCE: 15966-776CIP
; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/215,855
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/839,446
; PRIOR FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 32
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: POLYX
; OTHER INFORMATION: CG55655_02
US-09-898-570-32

Query Match 9.0%; Score 96; DB 10; Length 294;
Best Local Similarity 23.8%; Pred. No. 0.089;
Matches 44; Conservative 31; Mismatches 58; Indels 52; Gaps 10;
Qy 27 EEINKAIDAIAROSEID-PMKY-----PDHADKFERHVGIVDF-----KGE-70
Db 103 EXINSLNDLVKVEKSEVNGSPSSVYTRILKSQHAAMF-RHFOQIMFYNDITAAKQEK 161
Qy 71 ---LAMRNIEARG-----LKQMKRGDANKGEE-----GIVRAHL--LIGVHDDIVS 113
Db 162 CFTFLRQLEAVAGKEMSEEDVNDMLHOGKWEVFNESLLEINITYKAQISEIQRKHELVN 221

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7
US-10-024-955-7

Query Match 100.0%; Score 1068; DB 9; Length 213;
Best Local Similarity 100.0%; Pred. No. 1,5e-94;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAIEOSETIDPMKVPDHAKFER 60
DB 1 MKKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAIEOSETIDPMKVPDHAKFER 60

QY 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANYKGEIGYKAHLIGVHDDIVSMEXDLAY 120
DB 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANYKGEIGYKAHLIGVHDDIVSMEXDLAY 120

QY 121 KIGDLHPTTHVSDIDDFVALSLEISDGNITMTSFEVROFANVNHIGLSILDPPIRG 180
DB 121 KIGDLHPTTHVSDIDDFVALSLEISDGNITMTSFEVROFANVNHIGLSILDPPIRG 180

QY 181 VLSDVLTALFODTVRKEMTKVLAPAFKRELEKN 213
DB 181 VLSDVLTALFODTVRKEMTKVLAPAFKRELEKN 213

RESULT 2
US-10-024-955-2
Sequence 2, Application US/10024955
Patent No. US20020168373A1

GENERAL INFORMATION:
APPLICANT: WAYNE R. THOMAS AND KAW-YAN CHUA
TITLE OF INVENTION: Allergenic Proteins and Peptides From House Dust Mite and Uses Therefor

NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993

ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMT-032CP2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 215 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2
US-10-024-955-2

Query Match 90.1%; Score 962; DB 9; Length 215;
Best Local Similarity 85.9%; Pred. No. 1.9e-84;

Matches 183; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAIEOSETIDPMKVPDHAKFER 60
DB 1 MKKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAIEOSETIDPMKVPDHAKFER 60

QY 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANYKGEIGYKAHLIGVHDDIVSMEXDLAY 120
DB 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANYKGEIGYKAHLIGVHDDIVSMEXDLAY 120

QY 121 KIGDLHPTTHVSDIDDFVALSLEISDGNITMTSFEVROFANVNHIGLSILDPPIRG 180
DB 121 KIGDLHPTTHVSDIDDFVALSLEISDGNITMTSFEVROFANVNHIGLSILDPPIRG 180

QY 181 VLSDVLTALFODTVRKEMTKVLAPAFKRELEKN 213
DB 181 VLSDVLTALFODTVRKEMTKVLAPAFKRELEKN 213

RESULT 3
US-09-898-570-28
Sequence 28, Application US/09898570
Patent No. US20020123612A1

GENERAL INFORMATION:
APPLICANT: GERLACH, VALERIE L.
APPLICANT: ELLERMAN, KAREN
APPLICANT: MACDOUGALL, JOHN R.
APPLICANT: SMITHSON, GLENDA

TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

FILE REFERENCE: 15966-776CIP
CURRENT APPLICATION NUMBER: US/09/898,570
CURRENT FILING DATE: 2001-07-03

PRIOR APPLICATION NUMBER: 60/198,293
PRIOR FILING DATE: 2000-04-19

PRIOR APPLICATION NUMBER: 60/198,645
PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: 60/210,809
PRIOR FILING DATE: 2000-06-09

PRIOR APPLICATION NUMBER: 60/199,476
PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/200,025
PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/224,610
PRIOR FILING DATE: 2000-08-11

PRIOR APPLICATION NUMBER: 60/200,024
PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/199,880
PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/218,591
PRIOR FILING DATE: 2000-07-17

PRIOR APPLICATION NUMBER: 60/271,814
PRIOR FILING DATE: 2001-02-27

PRIOR APPLICATION NUMBER: 60/215,855
PRIOR FILING DATE: 2000-07-03

PRIOR APPLICATION NUMBER: 09/839,446
PRIOR FILING DATE: 2001-04-19

NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 28
LENGTH: 294
TYPE: PRT

ORGANISM: Unknown Organism

FEATURE:
OTHER INFORMATION: Description of Unknown Organism: POLYX
OTHER INFORMATION: h_nh0778p17_A

US-09-898-570-28

Query Match 9.0%; Score 96; DB 10; Length 294;
Best Local Similarity 23.8%; Pred. No. 0.089;
Matches 44; Conservative 31; Mismatches 58; Indels 52; Gaps 10;

QY 27 EEINKAIDDAIAAIEOSETID-PMKV-----PDHAKFERHVGIVDF-----KGE- 70

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 21, 2003, 10:09:25 ; Search time 31 Seconds
(without alignments)
213.480 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 1068
Sequence: 1 MMKFLIAAFAVAVSADPI.....VRKENTKVLAPAFKRELEKN 213

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PC6_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCUTS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1068	100.0	213	9	US-10-024-955-7
2	962	90.1	215	9	US-10-024-955-2
3	96	9.0	294	10	US-09-898-570-28
4	96	9.0	294	10	US-09-898-570-30
5	96	9.0	294	10	US-09-898-570-32
6	90	8.4	1786	9	US-09-742-096-3
7	86.5	8.1	660	10	US-09-815-242-5470
8	86.5	8.1	664	10	US-09-815-242-12179
9	83.5	7.8	412	9	US-09-738-626-3526
10	83	7.8	845	9	US-09-991-496-110
11	83	7.8	845	10	US-09-874-923-110
12	82	7.7	461	10	US-09-841-132-399
13	82	7.7	785	9	US-09-738-626-5445
14	80	7.5	831	9	US-09-738-626-5468
15	79.5	7.4	718	10	US-09-815-242-10901
16	79.5	7.4	6304	9	US-10-147-026-16
17	79	7.4	550	9	US-09-738-626-5843
18	78.5	7.4	810	10	US-09-815-242-5586
19	78.5	7.4	818	10	US-09-815-242-12424

20	78.5	7.4	818	10	US-09-815-242-12789	Sequence 12789, A
21	78.5	7.4	1209	9	US-09-738-626-4574	Sequence 4574, Ap
22	78	7.3	215	10	US-09-815-242-12956	Sequence 12956, A
23	78	7.3	215	10	US-09-815-242-13084	Sequence 13084, A
24	77.5	7.3	946	10	US-09-828-423-3	Sequence 3, Appl1
25	77	7.2	722	10	US-09-815-242-11000	Sequence 11000, A
26	76.5	7.2	161	10	US-09-815-242-4914	Sequence 4914, Ap
27	76.5	7.2	166	10	US-09-815-242-10614	Sequence 10614, A
28	76	7.1	308	9	US-09-738-626-3637	Sequence 3637, Ap
29	76	7.1	567	10	US-09-815-242-13511	Sequence 13511, A
30	75.5	7.1	905	10	US-09-905-983-58	Sequence 58, Appl
31	75	7.0	572	10	US-09-815-242-5526	Sequence 5526, Ap
32	75	7.0	572	10	US-09-815-242-12351	Sequence 12351, A
33	75	7.0	1044	10	US-09-728-137-2	Sequence 2, Appl1
34	74.5	7.0	285	9	US-09-738-626-3564	Sequence 3564, Ap
35	74.5	7.0	357	9	US-10-260-877-128	Sequence 128, App
36	74.5	7.0	357	10	US-09-815-242-10981	Sequence 10981, A
37	74	6.9	928	9	US-09-860-846-41	Sequence 41, Appl
38	74	6.9	928	10	US-09-861-289-41	Sequence 41, Appl
39	74	6.9	2437	10	US-09-815-242-5834	Sequence 5834, Ap
40	74	6.9	6281	10	US-09-815-242-12996	Sequence 12996, A
41	73.5	6.9	445	9	US-09-996-634-135	Sequence 135, App
42	73.5	6.9	666	10	US-09-815-242-13637	Sequence 13637, A
43	73.5	6.9	1263	9	US-09-738-626-5418	Sequence 5418, Ap
44	73.5	6.9	1480	9	US-09-568-756-2	Sequence 2, Appl1
45	73.5	6.9	1993	9	US-10-098-979-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-10-024-955-7
Sequence 7, Application US/10024955
Patent No. US20020168373A1
GENERAL INFORMATION:
APPLICANT: WAYNE R. THOMAS AND KAW-YAN CHUA
TITLE OF INVENTION: Allergenic Proteins and Peptides From House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,955
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMT-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
TOPOLOGY: linear

;; TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSA1 DISEASE VIRUS
;; TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
;; TITLE OF INVENTION: THEREON
;; NUMBER OF SEQUENCES: 15
;; CORRESPONDENCE ADDRESSES:
;; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;; STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
;; CITY: ARLINGTON
;; STATE: VIRGINIA
;; COUNTRY: USA
;; ZIP: 22202
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/031.655
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/219,262
;; FILING DATE: 29-MAR-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: OBLON, NORMAN F
;; REGISTRATION NUMBER: 24,618
;; REFERENCE/DOCKET NUMBER: 2/47-047-27
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 413-3000
;; TELEFAX: (703) 413-2220
;; TELEX: 248855 OPAT UR
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1012 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-09-031-655-14
;;
Query Match 7.1%; Score 75.5; DB 3; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13; Indels 21; Gaps 9;
Matches 36; Conservative 35; Mismatches 63;
QY 40 IEQSETIDPMKVPDHDADKFERHVGIVDFKELA-MRNIENGKLGKQKROGDANVKGEECI 98
DB 102 VRSRLVSSSTLPBGVYALNGTINAVTFQSGLSLTDVSYNGL---MSATANINDKIG- 156
QY 99 VKAHLIGVHDIYGM--EYDLAY-KLGDLPHTHVISDIQFVVALSLFIDEGNI-TM 154
DB 157 ---NVLVGGVTVLSLPSTYDGLGVRLGDPFPAIGL----DPKVVATCDSSDRPRVYTI 208
QY 155 TSFEVRFANVNVNHIGLSILDPFVGLSDVLTAI 189
DB 209 TAAIDNYQFSSQY-QTGGGVYI--TLFSANIDATITSL 240
;;
RESULT 50
US-08-433-522A-2
;; Sequence 2, Application US/08433522A
;; Patent No. 6013514
;; GENERAL INFORMATION:
;; APPLICANT: CHONG, Pele
;; APPLICANT: THOMAS, Wayne
;; APPLICANT: YANG, Yan Ping
;; APPLICANT: LOOSMORE, Sheena
;; APPLICANT: SIA, Dwo Yuan Charles
;; APPLICANT: KLEIN, Michel
;; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
;; NUMBER OF SEQUENCES: 55
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sim & McBurney
;; STREET: 6TH Floor, 330 University Avenue
;; CITY: Toronto

;; STATE: Ontario
;; COUNTRY: Canada
;; ZIP: M5G 1R7
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/433,522A
;; FILING DATE: 12-SEP-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEWART, Michael I
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-434 MIS-jb
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1153
;; TELEFAX: (416) 595-1155
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 797 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-08-433-522A-2
;;
Query Match 7.0%; Score 75; DB 3; Length 797;
Best Local Similarity 19.4%; Pred. No. 10;
Matches 62; Conservative 49; Mismatches 88; Indels 120; Gaps 16;
QY 2 MKFLIIAIAVAF---VAVSADP-----IHDKITEELNKAI-----DDAIAAI 40
DB 1 MKRLIIASLIFGTTTTFAPFAVAKDIRVGVGDLEQDIRASLPVRAGQRTVDNVANI 60
QY 41 EQSETI---DPMKVPDHDADK-----ERHVGIVDFK-----ELAMRNIEANGKQK 84
DB 61 VRSLEFVSGRFDVKAHQEGDVLVSVVAKSIISDVXIKGNSVIPTPEALKONIDANGFK-- 118
QY 85 KROGDANVKG-----EGIV-----KALLLIGVHDDIVSMXY 116
DB 119 --VGDLIREKLINEFAKSVKEHYASVGRYNAVTEPIVNTLPNNRAEILIIQINEDDKAKIA 176
QY 117 DLAYKLG-----DLHPTT-----HVISDIQDF-----V 139
DB 177 SLTFKKNESVSSSTIQEQMELQPDSSWMLKMGKFBGAGFEKDLQSLRDYLLNNGYAKAI 236
QY 140 VALSLEISDEG---NITMTSEVYRQF---ANVNVHIGLST-LDPIFGVLSDVLTAFQ 191
DB 237 TRTDVOLNDEKTRVNTTIDVNEGLOYDLRSARLIGNLGMASALEPPLSLALH-----LN 290
QY 192 DTVRKEMTFVLAAPAKREL 210
DB 291 DTFRRSDIADVENAKIAKL 309

Search completed: February 21, 2003, 10:10:05
Job time : 22 secs

[illegible]

```

      RESULT 45
US-09-031-655-1
: Sequence 1, Application US/09031655
Patent No. 6017759
: GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENGEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAI DISEASE VIRUS
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,655
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/219,262
FILING DATE: 29-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: GLS
US-09-031-655-1

Query Match 7.1%; Score 75.5; DB 3; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

```

```

Db      102  VSRSLVRSSTLPQGYVALNTIAAFQGSLSLSTDPVSYNGL-----MSATANIINDIG- 156
Qy      99  VKAHLILGHHDDISM--EKVDLAY-RKIGDLPHTTHVLSDDIQDFVALSLISPEGNI-TM 154
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157  ---NVLIGEGETVYSLSTPSTYDLGIVRKGRDPIPALG-----DKPMATCCSSDRPRVYTI 208
Qy      155  TSEFVEQFANVYNIIGLSILDPFEGVLSDYVTAI 189
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      209  TAAADYQFSSOY-OTGCVYTI--TLESANIDALISL 240

```

```

RESULT 46
US-09-031-655-2
: Sequence 2, Application US/09031655
Patent No. 6017759
GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENGEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAI DISEASE VIRUS
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
TITLE OF INVENTION: THERON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBOLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,655
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/219,262
FILING DATE: 29-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBOLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TEXT: 24885 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: Protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: DS326
US-09-031-655-2

Query Match          7.1%; Score 75.5; DB 3; Length 1012;
Best local similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

QY      40 IEQETIDPMKVPDHADFEHRHGVLDPEKGELA-MRNTIARGLKMKROGDANYKGEIGI 98
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      102 VSRSLTVRSSTLPGLGVALNGFINAVTFQGSLSLTLVDYNGL----MSAFNINDKIG- 156

OY      99 VKAHLLIGVHDIIYSM--EYDLAY-KLGDDHPHTTHIVISDIQDFVALSLEISDEINI-TM 154
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157 ---NLVLGGVGYTSLPSYDLGVYKRLGDPILPAIGL-----DKPVAVACDSSDRPRVYTI 208

```

APPLICATION NUMBER: US/08/219,262B
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: D78
US-08-219-262B-4

Query Match 7.1%; Score 75.5; DB 1; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

QY 40 IBSSEITDPMKVPDHDKFEFHVGIYDFGELA-MRNIERAGIKOKKRGDANVKEEGI 98
DB 102 VSRSLTVRSSTLPGGYALNGTINAVTFQGSLSLTDVSYNGL---MSATNINDKIG- 156
QY 99 VKAHLIGVHDDIVSM-EYDLAY-KLGDLPHTHYISDIODFVVALSLIEISDEGNI-TM 154
DB 157 ---NVLVGSGVYLSLPTSYDLGYRLGDPPIAIGL-----DPKMWATCDDSDRPVYTI 208
QY 155 TSEFVRQFANVNVNHIGLSILDPFICVLSVDVLTAI 189
DB 209 TAADYQFSSQY-QTGGVTI--TLFSANIDAITSL 240

RESULT 43
US-08-219-262B-12
Sequence 12, Application US/08219262B
Patent No. 5788970
GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENGEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,262B
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-219-262B-12

Query Match 7.1%; Score 75.5; DB 1; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

QY 40 IBSSEITDPMKVPDHDKFEFHVGIYDFGELA-MRNIERAGIKOKKRGDANVKEEGI 98
DB 102 VSRSLTVRSSTLPGGYALNGTINAVTFQGSLSLTDVSYNGL---MSATNINDKIG- 156
QY 99 VKAHLIGVHDDIVSM-EYDLAY-KLGDLPHTHYISDIODFVVALSLIEISDEGNI-TM 154
DB 157 ---NVLVGSGVYLSLPTSYDLGYRLGDPPIAIGL-----DPKMWATCDDSDRPVYTI 208
QY 155 TSEFVRQFANVNVNHIGLSILDPFICVLSVDVLTAI 189
DB 209 TAADYQFSSQY-QTGGVTI--TLFSANIDAITSL 240

RESULT 44
US-08-219-262B-14
Sequence 14, Application US/08219262B
Patent No. 5788970
GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENGEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,262B
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-219-262B-14

Query Match 7.1%; Score 75.5; DB 1; Length 1012;

TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
ON THE INVENTION: THEREON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,262B
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: GLS
US-08-219-262B-1

Query Match 7.1%; Score 75.5; DB 1; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

QY 40 IEOSETIDPMKVPDADKERRHVGIVDFKGLA-MRNIERAGLKQKRGDANVAGEEG 98
DB 102 VRSLSVRSSTLPGGYVALNGTINAVTFQGSLSLTDVSYNGL-----MSATNINDKIG- 156

QY 99 VAAHLIGVHDDIVSM--EYDLAY-KLGLDHPHTHVISDIOFVVALSLEISDEGNI-TM 154
DB 157 ---NVLVGGVTVLSLPTSYDGLGYVRLGDPFAIGL-----DPKMWATCDDSDRPVYTI 208

QY 155 TSFEVQFANVNVNHIGLSILDPFGVLSDLVLTAI 189
DB 209 TAADYQFSSQY-QTGGVTI--TLFSANIDAITSL 240

RESULT 41
US-08-219-262B-2
Sequence 2, Application US/08219262B
Patent No. 5788970
GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENDEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
ON THE INVENTION: THEREON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA

COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,262B
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: DS326
US-08-219-262B-2

Query Match 7.1%; Score 75.5; DB 1; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

QY 40 IEOSETIDPMKVPDADKERRHVGIVDFKGLA-MRNIERAGLKQKRGDANVAGEEG 98
DB 102 VRSLSVRSSTLPGGYVALNGTINAVTFQGSLSLTDVSYNGL-----MSATNINDKIG- 156

QY 99 VAAHLIGVHDDIVSM--EYDLAY-KLGLDHPHTHVISDIOFVVALSLEISDEGNI-TM 154
DB 157 ---NVLVGGVTVLSLPTSYDGLGYVRLGDPFAIGL-----DPKMWATCDDSDRPVYTI 208

QY 155 TSFEVQFANVNVNHIGLSILDPFGVLSDLVLTAI 189
DB 209 TAADYQFSSQY-QGVTI--TLFSANIDAITSL 240

RESULT 42
US-08-219-262B-4
Sequence 4, Application US/08219262B
Patent No. 5788970
GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENDEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
ON THE INVENTION: THEREON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

Query Match 7.2%; Score 77; DB 3; Length 1012;
Best Local Similarity 23.4%; Pred. No. 8.6;
Matches 39; Conservative 36; Mismatches 64; Indels 28; Gaps 10;

QY 40 IEQSEITIDPKVYDHDADKFERHGYIDFKGELA-MNIEARGLKOKKROGDANVKGEEGI 98
DB 102 VSRSLTVRSSTLPGYVALNGTINAVTFQGSISELTDVSYNGL---MSATANINDKIG- 156
QY 99 VKAHLIGVHDDIVSM--EYDLAY-KLGDLPHTHYISDIQDPVVALSLEISDEGNI-TM 154
DB 157 ---NLVAGGCVYVLSLPTSDGLGYVRLGDPIDPAIGL-----DPKMATCDSSDRPRVYTI 208
QY 155 TSFEVRQFANVNVHIGLSLIDPIFGVLSVLTAI 194
DB 209 TAADYQFSSQYOP-GGVTI--TLFSANIDAITSLSVGELVPTQTSV 252

RESULT 38
US-07-944-943-2
; Sequence 2, Application US/07944943
; Patent No. 5518724
; GENERAL INFORMATION:
; APPLICANT: SNYDER, DAVID B.
; APPLICANT: VAKHARIA, VIKRAM
; TITLE OF INVENTION: NOVEL INFECTIOUS BURSAL DISEASE VIRUS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/944,943
; FILING DATE: 19920915
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kelber, Steven B.
; REGISTRATION NUMBER: 30,073
; REFERENCE/DOCKET NUMBER: 2284-029-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1012 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-944-943-2

Query Match 7.1%; Score 75.5; DB 1; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

QY 40 IEQSEITIDPKVYDHDADKFERHGYIDFKGELA-MNIEARGLKOKKROGDANVKGEEGI 98
DB 102 VSRSLTVRSSTLPGYVALNGTINAVTFQGSISELTDVSYNGL---MSATANINDKIG- 156
QY 99 VKAHLIGVHDDIVSM--EYDLAY-KLGDLPHTHYISDIQDPVVALSLEISDEGNI-TM 154
DB 157 ---NLVAGGCVYVLSLPTSDGLGYVRLGDPIDPAIGL-----DPKMATCDSSDRPRVYTI 208
QY 155 TSFEVRQFANVNVHIGLSLIDPIFGVLSVLTAI 189
DB 209 TAADYQFSSQYOP-GGVTI--TLFSANIDAITSLSVGELVPTQTSV 252

DB 209 TAADYQFSSQYOP-GGVTI--TLFSANIDAITSL 240

RESULT 39
US-07-944-525-2
; Sequence 2, Application US/07944525
; Patent No. 5632989
; GENERAL INFORMATION:
; APPLICANT: SNYDER, DAVID B.
; APPLICANT: VAKHARIA, VIKRAM
; APPLICANT: LUETTICHEN, HEINRICH D.
; TITLE OF INVENTION: ATTENUATED, LIVE VACCINE FOR DELAWARE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER AND
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/944,525
; FILING DATE: 19920914
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KELBER, STEVEN B.
; REGISTRATION NUMBER: 30,073
; REFERENCE/DOCKET NUMBER: 2284-028-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1012 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-944-525-2

Query Match 7.1%; Score 75.5; DB 1; Length 1012;
Best Local Similarity 23.2%; Pred. No. 13;
Matches 36; Conservative 35; Mismatches 63; Indels 21; Gaps 9;

QY 40 IEQSEITIDPKVYDHDADKFERHGYIDFKGELA-MNIEARGLKOKKROGDANVKGEEGI 98
DB 102 VSRSLTVRSSTLPGYVALNGTINAVTFQGSISELTDVSYNGL---MSATANINDKIG- 156
QY 99 VKAHLIGVHDDIVSM--EYDLAY-KLGDLPHTHYISDIQDPVVALSLEISDEGNI-TM 154
DB 157 ---NLVAGGCVYVLSLPTSDGLGYVRLGDPIDPAIGL-----DPKMATCDSSDRPRVYTI 208
QY 155 TSFEVRQFANVNVHIGLSLIDPIFGVLSVLTAI 189
DB 209 TAADYQFSSQYOP-GGVTI--TLFSANIDAITSLSVGELVPTQTSV 252

RESULT 40
US-08-219-262B-1
; Sequence 1, Application US/08219262B
; Patent No. 5788970
; GENERAL INFORMATION:
; APPLICANT: VAKHARIA, VIKRAM
; APPLICANT: SNYDER, DAVID B.
; APPLICANT: MENGEL-WHERSAT, STEPHANIE A
; TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS


```

; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,774
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/074,579
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceirone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0505 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 946 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENEBAK
; CLONE: g133985
US-09-388-774-3

```

```

Query Match          7.3%; Score 77.5; DB 4; Length 946;
Best Local Similarity 19.8%; Pred. No. 6.8;
Matches 57; Conservative 46; Mismatches 84; Indels 101; Gaps 14;

```

```

QY 19 PIHYKTEIENKAIKD--DAIAIEOSEITIDPMKVPDHDADKFERHGVDFKGEI-AMRN 75
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 301 PNDLPRIKNIIFVLDVSSMNGVKMKQVLEAMKTIIDLRADHFSVIDFNONIRITWRN 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 -----IEANGKQMKRQGDANVKEGEGIVKAHLLIGVHDDIVSME---YDLAYK 121
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 361 DLFOLQKHLQIAKRYIEKIQPSGFTNI--NEALLRAIFILNEANNLGLDPSNSVLLIL 418
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 122 LDDHPTT-----HVISDIQDFVALSL-----EISDE-----G 150
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 419 VSDGPTVELKLSKIQKNVKNENIDNISLFSLGMGFVDVDFLKRLSNENHGIQRIYQ 478
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 151 NITMTSEFEYKOFANVY-----NHIGGLSL-----DPT- 178
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 479 N-QDITSSQALKRYNOVSTPLRNQGNPIPHTSVTDYTONNFHNYGSGSEIVAGKFDPAK 537
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 179 FGVLSDVLTAI-----FQDTVRKENTKVLAPAFKREL 210
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 538 LQJIESVITATSNQVLVETLAQMDLDQDFLSKD--KHADDPFTKRL 583
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

RESULT 31
US-09-425-453A-8
; Sequence 8, Application US/09425453A
; Patent No. 6468793
; GENERAL INFORMATION:
; APPLICANT: Teem, John L.
; TITLE OF INVENTION: CFTR Genes and Proteins for Cystic Fibrosis Gene Therapy
; FILE REFERENCE: FSU-99XC1
; CURRENT APPLICATION NUMBER: US/09/425,453A
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/105,444
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1480
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE

```

```

; LOCATION: (1)..(1480)
US-09-425-453A-8

```

```

Query Match          7.3%; Score 77.5; DB 4; Length 1480;
Best Local Similarity 28.5%; Pred. No. 13;
Matches 35; Conservative 21; Mismatches 50; Indels 17; Gaps 7;

```

```

QY 97 GIVKAHLLIGVHDDIVSMEDLAYKLGDLHPTTHVISPDIQDFVALSLSDSGNITMTS 156
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 500 GIKENIIFGV-----SYD-EYRYSYVAKQQLLEDISKFAEKDNM-VLGGGITTLSG 550
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 157 FEVRO--FANVYNHIGSLIDPIFGVLSVLT--AIFQDTVRKEM---TKVLAPAFKRE 209
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 551 GQMAKISLARAVYKADADLYLDSPPGYL-DVLTKEIFEESCVCYCKLANKTRILVTSKMEH 609
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 210 LEK 212
   | : |
DB 610 LKK 612

```

```

RESULT 32
US-09-425-453A-18
; Sequence 18, Application US/09425453A
; Patent No. 6468793
; GENERAL INFORMATION:
; APPLICANT: Teem, John L.
; TITLE OF INVENTION: CFTR Genes and Proteins for Cystic Fibrosis Gene Therapy
; FILE REFERENCE: FSU-99XC1
; CURRENT APPLICATION NUMBER: US/09/425,453A
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/105,444
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1480
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-425-453A-18

```

```

Query Match          7.3%; Score 77.5; DB 4; Length 1480;
Best Local Similarity 28.5%; Pred. No. 13;
Matches 35; Conservative 21; Mismatches 50; Indels 17; Gaps 7;

```

```

QY 97 GIVKAHLLIGVHDDIVSMEDLAYKLGDLHPTTHVISPDIQDFVALSLSDSGNITMTS 156
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 500 GIKENIIFGV-----SYD-EYRYSYVAKQQLLEDISKFAEKDNM-VLGGGITTLSG 550
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 157 FEVRO--FANVYNHIGSLIDPIFGVLSVLT--AIFQDTVRKEM---TKVLAPAFKRE 209
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 551 GQMAKISLARAVYKADADLYLDSPPGYL-DVLTKEIFEESCVCYCKLANKTRILVTSKMEH 609
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 210 LEK 212
   | : |
DB 610 LKK 612

```

```

RESULT 33
US-08-219-262B-3
; Sequence 3, Application US/08219262B
; Patent No. 5788970
; GENERAL INFORMATION:
; APPLICANT: VAKHARIA, VIKRAM
; APPLICANT: SNYDER, DAVID B
; APPLICANT: MENDEL-WHERSAT, STEPHANIE A
; TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
; TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
; TITLE OF INVENTION: THEREON
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON

```


COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC Compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,655
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/219,262
FILING DATE: 29-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: OHLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: STC
US-09-031-655-8

Query Match 7.3%; Score 78; DB 3; Length 1012;
Best Local Similarity 23.2%; Pred. No. 6.6;
Matches 39; Conservative 37; Mismatches 64; Indels 28; Gaps 10;

QY 40 IESETIDPKKVPDHDKFERHVGIVDFKGLA-MRIIEARGLKQKROGDANVKGEGE 98
DB 102 VSSLSVLRSTLRGGYALNGTINAVTFQGSLELDVSYNGL-----MSATAINDKIG- 156
QY 99 VKAHLIGVHDDIVSM--EYDLAY-KLGDLPHTTHVISDIODEVALSLEISDEGNI-TM 154
DB 157 ---NVLAGEGVTLSPYSDLGIVRLGDPPIAIGL-----DPKMAVTCDSRPARYTI 208
QY 155 TSEVFNQFANVNHIGLSILDPFGVLSVLA-----IFQDIYR 195
DB 209 TAADDYQFSSQYQ-PGGVTI--TLFSANIDATISLSVSGELVFQTSVQ 253

RESULT 29
US-09-074-579-3
Sequence 3, Application US/09074579
Patent No. 6001596
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Guegler, Karl J.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS

SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,579
FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Carione, Michael C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0505 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENE BANK
CLONE: 9133985
US-09-074-579-3

Query Match 7.3%; Score 77.5; DB 3; Length 946;
Best Local Similarity 19.8%; Pred. No. 6.8;
Matches 57; Conservative 46; Mismatches 84; Indels 101; Gaps 14;

QY 19 PIHYDITEERINKAID--DAIAIEQSETIDPKKVPDHDKFERHVGIVDFKGL-AMRN 75
DB 301 PDNLDPKPKNLFVIDYSSGMGVKKKQYEAANKTLLDLRADHFSVIDFNQINITWRN 360
QY 76 -----IARGLKQKROGDANVKGEGIVKHAHLIGVHDDIVSKE---YDLAYK 121
DB 361 DLFQLOKHRLQIAKRYIEKIQPSGCTNI--NEALRAIFLNEANNLGLDPNSVSLIIL 418
QY 122 LQDLHPPT-----HYSDIODEVALSL-----BISDE-----G 150
DB 419 VSDGDPFVGLKSLQKKNKENIDNISLSFGMGFVDYDFLKRLSNENHIGIAQRIYG 478
QY 151 NITMTSEVFNQFANV-----NHHIGLSIL-----DPI- 178
DB 479 N-QDTSSQLKRYNOVSTPLRNVOFNYPHTSVDTQNNFNHYPFGSGSELVYAGKRDPAK 537
QY 179 FGVLSVLA-----FQDYVREKMTKVLAPAFKREL 210
DB 538 LDQIESVITATSNATQVLVETTLAQMDLDQFLSKD--KHADPDTFKL 583

RESULT 30
US-09-388-774-3
Sequence 3, Application US/09388774
Patent No. 6228991
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Guegler, Karl J.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible

```

GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
TITLE OF INVENTION: MENDEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAI DISEASE VIRUS
TITLE OF INVENTION: CNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLOJ, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,262B
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLOJ, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursai disease virus
STRAIN: SRC
US-08-219-262B-8

Query Match          7.3%, Score 78; DB 1; Length 1012;
Best Local Similarity 23.2%, Pred. No. 6.6;
Matches 39; Conservative 37; Mismatches 64; Indels 28; Gaps 10;

QY      40 IEQSETIDPMVPPHADKFEKHGCVYDKRGELA-WRNIEARKLKMKGODNAYGGEET 98
Db       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      99 VKAHLLIGVHDIVSM--EYDLAY-KLGDLHTTVISDIODFYVALSLDISDEGNI-TM 154
Db       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157 ---NVLVGEGTYVLSPRYSYDLGYRLDDPIPAIDL-----DPKNVAITCDSSDRRRVYTI 208
QY      155 TSFEVRQAFNAVNNHIGLSILDDPIFGVLSDVLFIA-----IFQDTVR 195
Db       ! : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      209 TAADYGRSSQYQP-GGVTI--TFESANIDAITSLSVGELAVFGQSIVQ 253

RESULT 27
US-09-031-655-7
Sequence 7, Application US/09031655
Patent No. 6017759
GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENDEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAI DISEASE VIRUS
TITLE OF INVENTION: CNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
NUMBER OF SEQUENCES: 15

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,655
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/219,262
FILING DATE: 29-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: 52/70
US-09-031-655-7

Query Match 7.3%; Score 78; DB 3; Length 1012;
Best Local Similarity 23.2%; Pred. No. 6.6;
Matches 39; Conservative 37; Mismatches 64; Indels 28; Gaps 10;

QY 40 IEQSTIDPMKVPDHDADFERHVGIVDPKGLA-MRNIEARGLKMKROGDANVGEBCI 98
DB 102 VSRSLTVRSSTPLPGVVALNGTINAVTFQGSISELTDVSYNGL---MSATFINDKIG- 156
QY 99 VKAHLLIGVHDDIYSM--EYDIAV-KLGDHPHTTVISDIOFVALSLSEDSGNI-TM 154
DB 157 ---NLVGEGVTVSLPSTYDGLGYRLDPTPAICL-----DPKNVATCDSSDRPRVYTI 208
QY 155 TSFEVQGRANVNVNHIGSLIDPIFGVLSDVLT-----TFQDTRV 195
DB 209 TAADDYQSSQYQP-GGVYTI--TLESANIDATITSLSIGELVFGISVQ 253

RESULT 28
US-09-031-655-8
Sequence 8, Application US/09031655
Patent No. 6017759
GENERAL INFORMATION:
APPLICANT: VAKHARIA, VIKRAM
APPLICANT: SNYDER, DAVID B
APPLICANT: MENDEL-WHERSAT, STEPHANIE A
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA

DB 664 ELQDORFAGFGSSDGDYETIRKTMKELKNSFRPE 700

RESULT 24

US-09-621-855-2

; Sequence 2, Application US/09621855

; Patent No. 6346608

; GENERAL INFORMATION:

; APPLICANT: Jaworski, Deborah J.

; Mang, Min

; Shilling, Lisa K.

; Burnham, Martin

; Fosberry, Andrew

; Hodgson, John E.

; Lawlor, Elizabeth

; Rosenberg, Martin

; Ward, Judith

; TITLE OF INVENTION: Mech

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert, Price & Rhoads

; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19103-2793

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM compatible

; OPERATING SYSTEM: Windows 95

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/621,855

; FILING DATE: 24-Jul-2000

; CLASSIFICATION: <Unknown>

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 09/040,843

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Falk, Stephen F

; REGISTRATION NUMBER: 36,795

; REFERENCE/DOCKET NUMBER: GM10082

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-994-2488

; TELEFAX: 215-994-2222

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 866 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-621-855-2

Query Match 7.4%; Score 78.5; DB 4; Length 866;

Best Local Similarity 22.6%; Pred. No. 4.6;

Matches 49; Conservative 37; Mismatches 86; Indels 45; Gaps 11;

DB 18 DPHDKTEE-----INKAIDAIAIROSSEIDPMKVPDHADEKREHVGIVDFKGEIA 72

DB 504 DTLHERVIGCKDAVNSISAKVRRAAGLK-----DP-----KRPISFLIPGPTG 548

DB 73 MRRIE-ARGL-KOMKROGANVGE--EGYKAHL---LIGV-----HDDIVMEVDLAY 120

DB 549 VGTETLALRALAESMGDDAMIRVDSMESEKHAVALGAPGYVHDGGLTEKVR 608

DB 121 KLGLDLPRTTHVIDS---TODFVALSLSEISDEGNTMTSEFVROPANVY---NHIGL 172

DB 609 K-----PYGVILFDEIEKAHPVFNILLOVDGHLDTIKGRIVDERNTIIIMTSNVAQ 663

DB 173 SLDPILFGLSDVLAIFODTVRKEMTKVLAIPAFAKRE 209

DB 664 ELQDORFAGFGSSDGDYETIRKTMKELKNSFRPE 700

RESULT 25

US-08-219-262B-7

; Sequence 7, Application US/08219262B

; Patent No. 5788970

; GENERAL INFORMATION:

; APPLICANT: VAKHARIA, VIRRAM

; SNIDER, DAVID B

; APPLICANT: MENGEL-WHERSAT, STEPHANIE A

; TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS

; TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBILON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

; STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/219,262B

; FILING DATE: 29-MAR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: OBILON, NORMAN F

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 2747-047-27

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 413-3000

; TELEFAX: (703) 413-2220

; TELEX: 248855 OPAT UR

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1012 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Infectious bursal disease virus

; STRAIN: 52/70

US-08-219-262B-7

Query Match 7.3%; Score 78; DB 1; Length 1012;

Best Local Similarity 23.2%; Pred. No. 6.6;

Matches 39; Conservative 37; Mismatches 64; Indels 28; Gaps 10;

DB 40 IROSEITIDPMKVPDHADEKREHVGIVDFKGEIA-MNIEARGLKQKROGDANVKEEGI 98

DB 102 VNSSLTVRSSTLPGGYALNGTINAVTFQSSLSLIDVSYNGL-----MSATININDKIG- 156

DB 99 VKAHLIGVHDDIVSM--EYDLAY-KLGLDLPRTTHVIDSIODFVALSLSEISDEGNI-TM 154

DB 157 ---NVLVGGVYVLSLPTSVLDGYVRLGPPIPAIGL-----DPKMAVATCDSSDRPRVYTI 208

DB 155 TSFEVROPANVNHIGSLDLPILFVLSVLTGTA-----IFQDPIVR 195

DB 209 TAADVDYQFSQYOP--GGVYI--TLFSANIDATYISLSIGGELVFQTSVQ 253

RESULT 26

US-08-219-262B-8

; Sequence 8, Application US/08219262B

; Patent No. 5788970

ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-094-889-1

Query Match 7.4%; Score 79.5; DB 1; Length 906;
Best Local Similarity 19.4%; Pred. No. 3.8;
Matches 48; Conservative 47; Mismatches 82; Indels 71; Gaps 10;

18 DPH--YDKITE-----EINKAIDATA-----AIEOSETIDPMKVPDHDAD 56
DB 361 DPLNIAIDIDMTKTRDLRQLRKAVMDHISDFLETNVPLLYIEAKSGNEKEVEKVAQ 420
QY 57 KFERHVG-IVDFKGLAMRNIEARGLKOK-----RQ 87
DB 421 VREHANKLVEVANLACISINNEGVKLVMMAATQIDSLCPQVYINALLAARPOSKVAQ 480
QY 88 GPNVGESEGIYKAHLILGVHDIYSEMYDLAYKGLDHPHTTHVISDIDPFVALSLEIS 147
DB 481 DNMDDVFKDQMEKQVRLTEAVDITSVDDFLYS-----ENHILDDVKKCVIA--LQES 532
QY 148 DEGNITMTSEFVRQFANVYVNHIGGLSLDPIRGVLSDLTAIFQDTRKEMTKVLA----- 203
DB 533 DVDLTDRTAGAINGRAARVYIH-----INAMENYETGVYTERV-LEATRLSETVA 583

QY 204 PAFKRELE 211
DB 584 PRAFEQVE 591

RESULT 21
US-09-040-843-4
Sequence 4, Application US/09040843
Patent No. 6124119
GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah J.
APPLICANT: Wang, Min
APPLICANT: Shilling, Lisa K.
APPLICANT: Burnham, Martin
APPLICANT: Fosberry, Andrew
APPLICANT: Hodgson, John E.
APPLICANT: Lawlor, Elizabeth
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: Mech
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Fast-Seq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,843
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/057,535
FILING DATE: 29-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GMI0082
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 672 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-040-843-4

Query Match 7.4%; Score 78.5; DB 3; Length 672;
Best Local Similarity 22.6%; Pred. No. 3.1;
Matches 49; Conservative 37; Mismatches 86; Indels 45; Gaps 11;

18 DPHYDKITEE-----INKAIDAIAEIOSETIDPMKVPDHDKFERHVGIVDFGELA 72
DB 358 DTLHERVIGOKDAVNSISAVRRARAGLK-----D-----KRLGISTIFLGPFG 402
QY 73 MRNIE-ARGL-KOMKROGDANVKG-EGIYKAHL-----HDDIVSEMYDLAY 120
DB 403 VGTETELARALAESMGFDDAMIRVDMSEFMKHAVSRLVGAPRGYGHDDGGLTEKVR 462
QY 121 KGLDLHPTTHVISD-----IDPFVALSLEISDEGNITMTSEFVRQFANVY-----NHIGGL 172
DB 463 K-----PYSVILFDEIEKAHPVFNILLQVLDGHLJDTKGRTVDERNNITIMTSNVGAQ 517

Db 208 PBTGAVLESPTLLADKKISNIREMLPVLEAVANAGKPLLIADVEGEALATA----- 262
Qy 162 FANVNNHIGL-----SILDPFGVLSVDLTAIFODTVRKEMTKVLAPARKRELEK 212
Db 263 ---VNTIRGIYKVAANAKAFGC---DRKKALMDIATLTGTGTVISEIGMELEK 311

RESULT 18
US-09-570-778A-12
; Sequence 12, Application US/09570778A
; Patent No. 6468773
; GENERAL INFORMATION:
; APPLICANT: Trimbur, Donald E.
; APPLICANT: Whited, Gregory M.
; APPLICANT: Selfinova, Olga V.
; TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase
; FILE REFERENCE: GC580-2
; CURRENT APPLICATION NUMBER: US/09/570,778A
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: US 60/134,868
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Clostridium pasteurianum
US-09-570-778A-12

Query Match 7.6%; Score 81; DB 4; Length 385;
Best Local Similarity 23.8%; Pred. No. 0.69;
Matches 48; Conservative 39; Mismatches 71; Indels 44; Gaps 10;

Qy 2 MKFLIAANAFVAVSA-DPIHDKITEELINKA--IDDAIAIEQSETIDPMKVPDHAKF 58
Db 161 IEFVIVSWNRLPLVSIINDPLMIKKRPAGLTATGMDALTHAESVSKDANPTD----- 215
Qy 59 ERHVGVDPKGEELAMRNIE--ARGLKOMKROGDANVKGEGIVKAKHLILGVHDDIVSMEX 116
Db 216 -----ALAIQATILIANLRQAVALGE-NLEARENNAVYASLLGMAFNANLQY 263
Qy 117 --DLAVKLDLHPHTTVHSIDIDFVVALSLEISDEGNITMTSEVNRQAFANVNNHIG---- 170
Db 264 VHAMAHQGLGGLYDMAHGVAN-----AMLLPVERYNLISNP---KKFADIAEFMGENTE 314
Qy 171 GIST-----LDPTEGVLSVY 185
Db 315 GLSVMEAAEKALDAMFRLSKDV 336

RESULT 19
US-08-336-618-24
; Sequence 24, Application US/08336618
; Patent No. 5763590
; GENERAL INFORMATION:
; APPLICANT: Peattie, Debra A.
; APPLICANT: Harding, Matthew W.
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
; TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
; STREET: Two Mililita Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,618
; FILING DATE: 09-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/963,325
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/777,752
; FILING DATE: 11-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/
; FILING DATE: 09-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: VP191-06A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-618-24

Query Match 7.4%; Score 79.5; DB 1; Length 458;
Best Local Similarity 22.1%; Pred. No. 1.3;
Matches 27; Conservative 29; Mismatches 49; Indels 17; Gaps 3;

Qy 18 DPIHDKITEELINKADDDIAIEQSETI-----DPRKVPDHD-KFERIV 62
Db 75 DKFSFDLGEVIAKMDIAVATMKVGEICRTCKPEYAVGSAGSPKIPPMATLVFE--V 132
Qy 63 GIVDFKGEELAMRNIEARGLKOMKROGDANVKGEGIVKAKHLILGVHDDIVSMEXDLYK 122
Db 133 ELFEKFGEDLTDDDEGILIRIRTRKGEYARNDGAIYVALGEGYKDKRLFQRELRFV 192
Qy 123 GD 124
Db 193 GE 194

RESULT 20
US-08-094-889-1
; Sequence 1, Application US/08094889
; Patent No. 5470966
; GENERAL INFORMATION:
; APPLICANT: Shiji HIRANO et al.
; TITLE OF INVENTION: NEURAL --CATENIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/094,889
; FILING DATE: July 22, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

QY 127 PTHVHISDIODEFVALSLEISD-----EGNITWTSFEVRQ 161
| | | | | : : : : : | | : : : | | | | |
Db 208 PETGAVALESPFLILADKRTISNIREMLPYLEAVAKAKPELLIIAEDVEGALATA----- 262

QY 162 FANVNHIGL-----SILDPFGVLSDLVLAIFQDTVRKEMTKVLAAPAKRELEK 212
| | | | | : : : : : | | : : : | | | | |
Db 263 ---VVNTIRGIYKVAAVKAPGFG---DRRKAMQDIATILTGCVISEIGMELEK 311

RESULT 16

US-08-432-697-32
; Sequence 32, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495, 0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-432-697-32

Query Match 7.6%; Score 81.5; DB 4; Length 548;

Best Local Similarity 20.4%; Pred. No. 1;

Matches 48; Conservative 37; Mismatches 85; Indels 65; Gaps 10;

QY 13 VAVSADPIHYDKTEIRINKAIDAI-AAIBQSETIDPMKVPPDHADKFEFRHGVDPKGL 71
| | | | | : : : : : | | : : : | | | | |
Db 107 VAAGNMPP-----DLKRGIDKAVTAAYVE--LKALSPCSDSKAIAOVGTISANSDE 156
QY 72 AMRNIEARGIKOKKRGDANVKGEGEIVAKAHLILGVHD--IVSMEDYLAAYKLGDL--H 126
| | | | | : : : : : | | : : : | | | | |
Db 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLQDELVDVEGQDFRGYLSPFYINK 207
QY 127 PTHVHISDIODEFVALSLEISD-----EGNITWTSFEVRQ 161
| | | | | : : : : : | | : : : | | | | |
Db 208 PETGAVALESPFLILADKRTISNIREMLPYLEAVAKAKPELLIIAEDVEGALATA----- 262

QY 162 FANVNHIGL-----SILDPFGVLSDLVLAIFQDTVRKEMTKVLAAPAKRELEK 212
| | | | | : : : : : | | : : : | | | | |
Db 263 ---VVNTIRGIYKVAAVKAPGFG---DRRKAMQDIATILTGCVISEIGMELEK 311

RESULT 17
US-08-466-248-32
; Sequence 32, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495, 0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-248-32

Query Match 7.6%; Score 81.5; DB 4; Length 548;

Best Local Similarity 20.4%; Pred. No. 1;

Matches 48; Conservative 37; Mismatches 85; Indels 65; Gaps 10;

QY 13 VAVSADPIHYDKTEIRINKAIDAI-AAIBQSETIDPMKVPPDHADKFEFRHGVDPKGL 71
| | | | | : : : : : | | : : : | | | | |
Db 107 VAAGNMPP-----DLKRGIDKAVTAAYVE--LKALSPCSDSKAIAOVGTISANSDE 156
QY 72 AMRNIEARGIKOKKRGDANVKGEGEIVAKAHLILGVHD--IVSMEDYLAAYKLGDL--H 126
| | | | | : : : : : | | : : : | | | | |
Db 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLQDELVDVEGQDFRGYLSPFYINK 207
QY 127 PTHVHISDIODEFVALSLEISD-----EGNITWTSFEVRQ 161
| | | | | : : : : : | | : : : | | | | |

APPLICATION NUMBER: US 08/073,381
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Graaahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH18-08AFA2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 547 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-06362-2

Query Match 7.6%; Score 81.5; DB 5; Length 547;
Best Local Similarity 20.4%; Pred. No. 1;
Matches 48; Conservative 37; Mismatches 85; Indels 65; Gaps 10;

QY 13 VAVSADPIHYDKITEIRINKAIDDAI-AAIEQSETIDPMKVPDPAHDKFERHVGIVDFKGL 71
DB 107 VAAGNPM-----DLKRGIDKAVTAAYEE---LKALSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIEARGLKOMKRGDANVKGEGIVKAHLIGVHD--IVSMEDLAYKLGDL--H 126
DB 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLDELVDVEGQDFRGYLSPEYFINK 207
QY 127 PTHVTSIDQDFVVALSLEISD-----EGNITMTSEVRO 161
DB 208 PETGAVELESPPILLADKKIISNIREMLPYLEAVAKAGKPLLIADVEBEALATA----- 262
QY 162 FANVNHIGL-----SIIDPIFGVLSVDYTAIFODTVRKEMTKVLAPARKRELEK 212
DB 263 ---VNTIRIGIVKVAVKAPGFG---DRKKAMLDIATLTGSGTVISEIGMELEK 311

RESULT 14
US-08-467-822-32
Sequence 32, Application US/08467822
Patent No. 5843460

GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauterbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NOCLEIC ACID SEQUENCES ENCODING SAID
NUMBER OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESS: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-822-32

Query Match 7.6%; Score 81.5; DB 2; Length 548;
Best Local Similarity 20.4%; Pred. No. 1;
Matches 48; Conservative 37; Mismatches 85; Indels 65; Gaps 10;

QY 13 VAVSADPIHYDKITEIRINKAIDDAI-AAIEQSETIDPMKVPDPAHDKFERHVGIVDFKGL 71
DB 107 VAAGNPM-----DLKRGIDKAVTAAYEE---LKALSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIEARGLKOMKRGDANVKGEGIVKAHLIGVHD--IVSMEDLAYKLGDL--H 126
DB 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLDELVDVEGQDFRGYLSPEYFINK 207
QY 127 PTHVTSIDQDFVVALSLEISD-----EGNITMTSEVRO 161
DB 208 PETGAVELESPPILLADKKIISNIREMLPYLEAVAKAGKPLLIADVEBEALATA----- 262
QY 162 FANVNHIGL-----SIIDPIFGVLSVDYTAIFODTVRKEMTKVLAPARKRELEK 212
DB 263 ---VNTIRIGIVKVAVKAPGFG---DRKKAMLDIATLTGSGTVISEIGMELEK 311

RESULT 15

US-09-472-971-3
Sequence 3, Application US/09472971
Patent No. 6197547
GENERAL INFORMATION:
APPLICANT: SOGO, Kazuyo
APPLICANT: YAMAGI, Hideki
APPLICANT: YURA, Takashi
TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
FILE REFERENCE: 1422-409P
CURRENT APPLICATION NUMBER: US/09/472,971
CURRENT FILING DATE: 1999-12-28
EARLIER APPLICATION NUMBER: JP10-372965
EARLIER FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO: 3
LENGTH: 548
TYPE: PRT
ORGANISM: Escherichia coli
US-09-472-971-3

Query Match 7.6%; Score 81.5; DB 4; Length 548;
Best Local Similarity 20.4%; Pred. No. 1;
Matches 48; Conservative 37; Mismatches 85; Indels 65; Gaps 10;

QY 13 VAVSADPIHYDKITEIRINKAIDDAI-AAIEQSETIDPMKVPDPAHDKFERHVGIVDFKGL 71
DB 107 VAAGNPM-----DLKRGIDKAVTAAYEE---LKALSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIEARGLKOMKRGDANVKGEGIVKAHLIGVHD--IVSMEDLAYKLGDL--H 126
DB 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLDELVDVEGQDFRGYLSPEYFINK 207

FILING DATE: 15-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/207,298
FILING DATE: 15-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/02619
FILING DATE: 15-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI88-08AF4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 547 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-461-722-2

Query Match 7.6%; Score 81.5; DB 4; Length 547;
Best Local Similarity 20.4%; Pred. No. 1;
Matches 48; Conservative 37; Mismatches 85; Indels 65; Gaps 10;

QY 13 VAVSADPIHYDKITEINKAIDDAI-AAIEOSETIDPMKVPDPHADKFEHVGIVDFKGL 71
DB 107 VAAGNPM-----DLRGIDKAVTAVEE---LKALSVPCSDSKALAQVETISANSD 156
QY 72 AMRNIEARGLKOMKROGDANVKEGEGIVKAHLIGVHD--TYSMEYDLAYKLGDL--H 126
DB 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLDELVDVEGMOFGRGYLSPYFINK 207
QY 127 PTHYISDIQDFVALSLEISD-----EGNITMTSEVRO 161
DB 208 PETGAVELESPITLADKKTISIREMLPYLEAVAKAPLLIIAEDVEGEALATA----- 262
QY 162 FANVNVHIGL-----SILDPPIRGVSLDVLTAFODTVRKEMTKVLAPAFKRELEK 212
DB 263 ---VNTIRGIYKVAAVKAPGFG---DRRKAMLDIATLTGTGTIVSEELGMLEK 311

RESULT 12
US-08-336-251-2
Sequence 2, Application US/08336251
Patent No. 6338952
GENERAL INFORMATION:
APPLICANT: Young, Richard S.
TITLE OF INVENTION: Stress Proteins and Uses Therefor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,251
FILING DATE: 03-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06362
FILING DATE: 06-JUN-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,381
FILING DATE: 04-JUN-1993

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/804,632
FILING DATE: 09-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/366,581
FILING DATE: 15-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/207,298
FILING DATE: 15-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/02619
FILING DATE: 15-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI88-08FA3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 547 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-251-2

Query Match 7.6%; Score 81.5; DB 4; Length 547;
Best Local Similarity 20.4%; Pred. No. 1;
Matches 48; Conservative 37; Mismatches 85; Indels 65; Gaps 10;

QY 13 VAVSADPIHYDKITEINKAIDDAI-AAIEOSETIDPMKVPDPHADKFEHVGIVDFKGL 71
DB 107 VAAGNPM-----DLRGIDKAVTAVEE---LKALSVPCSDSKALAQVETISANSD 156
QY 72 AMRNIEARGLKOMKROGDANVKEGEGIVKAHLIGVHD--TYSMEYDLAYKLGDL--H 126
DB 157 TVGKLIAEAMDKV-----GKEGVITVEDGTGLDELVDVEGMOFGRGYLSPYFINK 207
QY 127 PTHYISDIQDFVALSLEISD-----EGNITMTSEVRO 161
DB 208 PETGAVELESPITLADKKTISIREMLPYLEAVAKAPLLIIAEDVEGEALATA----- 262
QY 162 FANVNVHIGL-----SILDPPIRGVSLDVLTAFODTVRKEMTKVLAPAFKRELEK 212
DB 263 ---VNTIRGIYKVAAVKAPGFG---DRRKAMLDIATLTGTGTIVSEELGMLEK 311

RESULT 13
PCT-US94-06362-2
Sequence 2, Application PC/TUS9406362
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Stress Proteins and Uses Therefor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06362
FILING DATE: 06-JUN-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:

Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DRUILHE, PIERRE
APPLICANT: DAUBESIES, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1786
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match 8.4%; Score 90; DB 4; Length 1786;
Best Local Similarity 23.3%; Pred. No. 0.68;
Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;

QY 22 YDKITEINKAIDDAIAIEQSEFIDPMKVPDHDKFERHVGIVDKGELAMENIEARGL 81
DB 979 FNVLDKVEETVISESELENNE---MDKAPFSEIFDNVKGIOENLIGMFSISIFSTIV 1034
QY 82 KMKRQGDANVKGEGIVAKHLLIGVHDDIVSMEDYLAAYKLGDLHPHTTHVSDIOEFVVA 141
DB 1035 IOSEKRVDLN---ENVVSSIL-----DNLENKKEGLKLNLENISSTEGVQETVTEHV-- 1083
QY 142 LSEIIDEQGNITMTSEV---RQFANVANHIGLS-----ILDPIGVLSDVUTA--IRQ 191
DB 1084 -----EQNV-YVDVDPAMKQDFLIGLIEAGLKEFMFNEDVKSSEDVYTVVEIKD 1135
QY 192 DTVRKEMTKVLAPAFKRELEKN 213
DB 1136 EPVQKEVEKEVSIIE-EMEN 1156

RESULT 5
US-08-336-618-12
Sequence 12, Application US/08336618
Patent No. 5763590
GENERAL INFORMATION:
APPLICANT: Peattie, Debra A.
APPLICANT: Harding, Matthew W.
APPLICANT: Livingston, David J.
TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,618
FILING DATE: 09-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/963,325

FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/777,752
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/
FILING DATE: 09-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: VPI91-06A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 459 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-618-12

Query Match 8.3%; Score 88.5; DB 1; Length 459;
Best Local Similarity 24.6%; Pred. No. 0.13;
Matches 30; Conservative 25; Mismatches 50; Indels 17; Gaps 3;

QY 18 DPHYDKITEINKAIDDAIAIEQSEFI-----DPMKVPDHD-KFERHV 62
DB 75 DKSFGLGGEVTKAMDIAIAIKVGEVCHITCKPRYAVGSAGSPKPIPNATLVE--V 132
QY 63 GIVDFKSELAMRNIENARGKMKRQGDANVKGEGIVAKHLLIGVHDDIVSMEDYLAAYKL 122
DB 133 ELFEFGEDLTEREDGIIIRIOTRGEYAKPRGAIIVEALEGYKDKLFDQRELRFEL 192
QY 123 GD 124
DB 193 GE 194

RESULT 6
US-08-336-618-26
Sequence 26, Application US/08336618
Patent No. 5763590
GENERAL INFORMATION:
APPLICANT: Peattie, Debra A.
APPLICANT: Harding, Matthew W.
APPLICANT: Livingston, David J.
TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,618
FILING DATE: 09-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/963,325
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/777,752
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/

OY	61	HVGJVDKGETLAMPNIEIRGJKOMRGODNAVNGEETIVANHLIGVHDIYSMEVDLAY	120
Db	61	HVGJVDEFGELAMPNIEIRGJKOMRGODNAVNGEETIVANHLIGVHDIYSMEVDLAY	120
OY	121	KLGDLHPHTNHSIODPVALSLEISDEGNITWTSPEVFOPANVVHIGLSILDPFEG	180
Db	121	KLGDLHPHTNHSIODPVALSLEISDEGNITWTSPEVFOPANVVHIGLSILDPFEG	180
OY	181	VLSDVLTAFI FODTVRKEMTKVLAPAFKRELEKN	213
Db	181	VLSDVLTAFI FODTVRKEMTKVLAPAFKRELEKN	213

```

US-08-553-336A-2
Sequence 2, Application US/08553336A
Patent No. 6413738
GENERAL INFORMATION:
APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540
FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38, 872
REFERENCE/DOCKET NUMBER: IMT-032CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 215 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-553-336A-2

```

Query Match	90.1%	Score 962	DB 4	Length 215
Best Local Similarity	85.9%	Pred. No. 1	1e-101	
Matches 183	Conservative 18	Mismatches 12	Indels 0	Gaps 0

[illegible]

```
QY 181 VLSVDLTAIFQDTRVKEMTKVLAPAFKRELEKN 2133
      |||||:|||||:|||||
Db 181 VLSVDLTAIFQDTRVRAEMTKVLAPAFKKELEKN 2133
```

RESULT 3
 US-08-462-778-2
 Sequence 2, Application US/08462778
 Patent No. 6077517
 GENERAL INFORMATION:
 APPLICANT: Thomas, Wayne R.
 APPLICANT: Chu, Kaw-tan
 TITLE OF INVENTION: Allergenic Protein and Peptides From
 TITLE OF INVENTION: House Dust Mite and Uses Therefor
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Lathive & Cockfield
 STREET: 60 State Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/462.778
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE: 12 March 1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Amy E. Mandragouras
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-227-7400
 TELEFAX: 617-227-5941
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 215 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-462-778-2

Query Match	89.8%	Score 959	DB 3	length 215
Best Local Similarity	85.4%	Pred. No. 2.4e-101		
Matches 182	Conservative 19	Mismatches 12	Indels 0	Gaps 0

[illegible]

RESULT 4
US-08-973-462-8
; Sequence 8, Application US/08973462B

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:06:55 ; Search time 14 seconds

(without alignments)
447.649 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068
Sequence: 1 MMKFLIAAFAVAVSADPI.....VRKEMTVLAPAKRELEKN 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6ackfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1068	100.0	213	4	US-08-553-336A-7
2	962	90.1	215	4	US-08-553-336A-2
3	959	89.8	215	4	US-08-462-778-2
4	90	8.4	1786	4	US-08-973-462-8
5	88.5	8.3	459	1	US-08-336-618-12
6	88.5	8.3	459	1	US-08-336-618-26
7	84.5	7.9	294	1	US-08-137-175A-9
8	84.5	7.9	294	4	US-08-479-017-9
9	84.5	7.9	823	4	US-09-134-001C-4081
10	82	7.7	362	4	US-09-134-001C-4057
11	81.5	7.6	547	4	US-08-461-722-2
12	81.5	7.6	547	4	US-08-336-251-2
13	81.5	7.6	547	5	PCT-US94-06362-2
14	81.5	7.6	548	2	US-08-467-822-32
15	81.5	7.6	548	4	US-09-472-971-3
16	81.5	7.6	548	4	US-08-432-697-32
17	81.5	7.6	548	4	US-08-466-248-32
18	81	7.6	385	4	US-09-570-778A-12
19	79.5	7.4	458	1	US-08-336-618-24
20	79.5	7.4	906	1	US-08-094-889-1
21	78.5	7.4	672	3	US-09-040-843-4
22	78.5	7.4	672	4	US-09-621-855-4
23	78.5	7.4	866	3	US-09-040-843-2
24	78.5	7.4	866	4	US-09-621-855-2
25	78	7.3	1012	1	US-08-219-262B-7
26	78	7.3	1012	1	US-08-219-262B-8
27	78	7.3	1012	3	US-09-031-655-7

28	78	7.3	1012	3	US-09-031-655-8	Sequence 8, Appl1
29	77.5	7.3	946	4	US-09-074-579-3	Sequence 3, Appl1
30	77.5	7.3	946	4	US-09-388-774-3	Sequence 3, Appl1
31	77.5	7.3	1480	4	US-09-425-453A-8	Sequence 8, Appl1
32	77.5	7.3	1480	4	US-09-425-453A-18	Sequence 18, Appl1
33	77	7.2	1012	1	US-08-219-262B-3	Sequence 3, Appl1
34	77	7.2	1012	1	US-08-219-262B-5	Sequence 5, Appl1
35	77	7.2	1012	2	US-08-708-541A-30	Sequence 30, Appl1
36	77	7.2	1012	3	US-09-031-655-3	Sequence 3, Appl1
37	77	7.2	1012	3	US-09-031-655-5	Sequence 5, Appl1
38	75.5	7.1	1012	1	US-07-944-525-2	Sequence 2, Appl1
39	75.5	7.1	1012	1	US-07-944-525-2	Sequence 2, Appl1
40	75.5	7.1	1012	1	US-08-219-262B-1	Sequence 1, Appl1
41	75.5	7.1	1012	1	US-08-219-262B-2	Sequence 2, Appl1
42	75.5	7.1	1012	1	US-08-219-262B-4	Sequence 4, Appl1
43	75.5	7.1	1012	1	US-08-219-262B-12	Sequence 12, Appl1
44	75.5	7.1	1012	1	US-08-219-262B-14	Sequence 14, Appl1
45	75.5	7.1	1012	3	US-09-031-655-1	Sequence 1, Appl1

ALIGNMENTS

```
RESULT 1
US-08-553-336A-7
; Sequence 7, Application US/08553336A
; Patent No. 6413738
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From
; NUMBER OF INVENTIONS: House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38, 872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)742-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-553-336A-7

Query Match      100.0%; Score 1068; DB 4; Length 213;
Best Local Similarity 100.0%; Pred. No. 9,1e-114;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MMKFLIAAFAVAVSADPIHYDKITEINKAIDAIAIEQSETIDPKVPDHDADKFER 60
Db      1 MMKFLIAAFAVAVSADPIHYDKITEINKAIDAIAIEQSETIDPKVPDHDADKFER 60
```


Db 78 G-----KGHSFYFSMDIREPOLPSSQISMVPMVGLKFSISNANIKISGKKAKOERFLKMSG 133
QY 108 HDDI-----VSMEDLAYKLGDLHPPTHVTSIDIQDFVVALSLEISDEGNITWTSPFEVRQFA 163
Db 134 NFDLSTEGHSISADL--KLGs-NPTS-----GKPTITCSCSSHI 170
QY 164 NVVN-HIGGLSI-----LDPIFGVLSVDLTAIFODTVRKEMTKVLAFAFK 207
Db 171 NSVHVHISGSDVGLQLGFHNOIESALQNVMSQVCENTVNSVSDLPQYFQ 222

RESULT 50

AAV36857
ID AAV36857 standard; Protein; 530 AA.

AAV36857;

07-OCT-1999 (first entry)

Protein involved in intermediate metabolism of nucleic acids.

Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
paratrachoma; inclusion conjunctivitis; genital disease; perithenitis;
nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
bartholinitis; pneumonia; venereal lymphogranulomatosis.

Chlamydia trachomatis.

W09928475-A2.

10-JUN-1999.

27-NOV-1998; 98WO-1B01939.

04-NOV-1998; 98US-0107077.

28-NOV-1997; 97ER-0015041.

17-DEC-1997; 97ER-0016034.

(GEST) GENSET.

Griffiths R;

WPI: 1999-371125/31.

Genome sequence of Chlamydia trachomatis

Disclosure: Page 731; 1755pp; English.
AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome
of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
can also be used to control growth of the microorganism. Chlamydia
trachomatis is responsible for a large number of diseases, e.g. eye
diseases such as conventional trachoma, nonendemic trachoma,
paratrachoma, and inclusion conjunctivitis; genital diseases such as
nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
perithenitis, bartholinitis; pneumonia; venereal lymphogranulomatosis,
and venereal lymphogranulomatosis. The polypeptides of the invention
may be of use in treating these diseases.

Sequence 530 AA;

Query Match 7.7%; Score 82; DB 20; Length 530;

Best Local Similarity 21.6%; Pred. No. 6.2; Mismatches 45; Conservative 32; Mismatches 73; Indels 58; Gaps 9;

QY 18 DPHVYKITEEI-NKAIDAIAIEQSEITDPKVPDHDK----FERHGIYDFKGEA 72

Db 3 EPLYLORISYPLYNKSIHN-SHPMSITSPPLEVSULTDSIKNLEKFLRVVKGELS 61

QY 73 MRNIEARGLKOMKROGDANVKGEGIVKAKHLILGVHDD-----IYSMEYDLAYKL 122

Db 62 NVSLQTS-----HLVFAIKDSKAVLNGAEFHRRSKYFDRKPKD 100
QY 123 GD---LHPTTHVTSIDIQDFVVALSLEISDEGNITWTSPFEVRQFANVNHIGLSTIDP- 177
Db 101 GDVYLHGKLVYAPRQYQIYAVALTFSGEGNL-----LOQFEERKQRLAEGYFDRK 154
QY 178 -----IFGVLSVDLTAIFODTVR 195
Db 155 RKKPLPSGARVIGVITSPPTGAVIQDILR 182

Search completed: February 21, 2003, 10:07:15
Job time : 44 secs

PF 20-JUL-2001; 2001WO-US23121.
 XX
 XX 20-JUL-2000; 2000US-0620412.
 PR 23-APR-2001; 2001US-0841132.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX Fling SP, Skeiky YAM, Probst P, Bhatia A;
 DR WPI: 2002-179901/23.
 XX
 PT Novel compositions comprising Chlamydia Cap1 protein and its use in the
 XX treatment of Chlamydia infection -
 PS Disclosure; Page 361-363; 537pp; English.
 XX
 CC The present invention describes compositions comprising a Chlamydia Cap1
 CC protein and methods for the diagnosis and therapy of Chlamydia infection.
 CC Chlamydia DNA and protein sequences from the present invention can have
 CC antibacterial and immunostimulant activities, and can be used in
 CC vaccines. Compounds from the present invention can be used for eliciting
 CC an immune response, specifically stimulating a Chlamydia-specific T-cell
 CC response or inhibiting the development of a Chlamydia infection in an
 CC animal. Methods from the present invention can be used; for detecting the
 CC presence of Chlamydia in a patient; to stimulate and/or expand T cells
 CC specific for a Chlamydia protein; and for treatment of a Chlamydia
 CC infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 461 AA;
 Query Match 7.7%; Score 82; DB 23; Length 461;
 Best Local Similarity 20.4%; Pred. No. 5.1;
 Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;
 QY 23 DXITEE---INKAIDDAIAEQSETIDPMKVPDHADKFE-----RH 61
 DB 219 NFKTKGIRILTKA---SISAIEESQNOVRIYVNDVESEYDLVAIGROFNATSIGLDN 275
 QY 62 VGIV-DFKGLAMRNTEANGKQMKRQD-----ANKGEGCIYKAHLILGVHDIYS 113
 DB 276 AGVIRDDRQVIVDETRMNPVITAIGDITGKWLAAHSAHSGVTAANKNISGHHB---V 332
 QY 114 MEYDLAYKGLDHPPTHVNSIDQDFVALSLSEISDEGNT-TWTSPEVRQ----- 161
 DB 333 MDYSALPVSIFTHP-----EIAWVGLSLQEAEOQNLPKATLTKRPFKAIGKAVAGAS 384
 QY 162 --FANVNH-----IGGLSILDPFEGVLSVDVLTATFQDTVRKEMT 199
 DB 385 DGFALVSHETQOILGAVYIGPHASSLIGEMTL---AIRNELT 425
 RESULT 49
 AAWI6827
 ID AAWI6827 standard; protein; 487 AA.
 AC AAWI6827;
 XX 28-JUL-1997 (first entry)
 DT
 XX Recombinant endotoxin neutralising polypeptide NCY139.
 DE
 XX Recombinant; endotoxin; neutralisation; RENP; NCY139;
 KM lipopolysaccharide; LPS; bactericidal;
 KM permeability; increase; BPT; detection; Gram-negative; bacterium;
 KM infection; prevention; treatment; disorder; mediation; stimulation;
 KM neutrophil; mononuclear cell; inhibition; shock; anaemia;
 KM disseminated intravascular coagulation; thrombocytopenia;
 KM thrombocytopenia; adult respiratory distress syndrome; ARDS;
 KM renal failure; liver disease; half life; human.
 XX
 OS Homo sapiens.
 OS Synthetic.

XX
 FH Key Location/Qualifiers
 FT Peptide 1..31
 FT /label= sig_peptide
 FT Misc-difference 179
 FT /note= "wild type Lys replaced by Gly"
 FT Misc-difference 181
 FT /note= "wild type Lys replaced by Asp"
 FT Misc-difference 191
 FT /note= "wild type Lys replaced by Asn"
 FT Misc-difference 192
 FT /note= "wild type Lys replaced by Gln"
 FT Misc-difference 198
 FT /note= "wild type Arg replaced by Gln"
 FT Misc-difference 200
 FT /note= "wild type Lys replaced by Val"
 FT Misc-difference 208
 FT /note= "wild type Lys replaced by Met"
 FT Misc-difference 216
 FT /note= "wild type Lys replaced by Asp"
 FT Misc-difference 229
 FT /note= "wild type Lys replaced by Glu"
 XX
 PN W09634873-A1.
 XX
 PD 07-NOV-1996.
 XX
 PP 01-MAY-1996; 96WO-US06134.
 XX
 PR 01-MAY-1995; 95US-0431517.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Marra NM, Scott RW;
 XX
 DR WPI: 1996-506090/50.
 XX
 XX Lipo:polysaccharide binding and endotoxin neutralising polypeptide
 PT - useful in diagnosis, prevention and treatment of Gram-negative
 PT bacterial infection and associated disorders and conditions
 PS
 PS Example 1; Page -: 147pp; English.
 XX
 CC The present sequence is the recombinant endotoxin neutralising
 CC polypeptide (RENP) B(CAT9) (NCY139), which comprises the human
 CC bactericidal/permeability increasing protein (bPIL) mutant
 CC Lys148Gly, Lys150Asp, Lys160Asp, Lys161Gln, Arg167Gln, Lys169Val,
 CC Lys177Met, Lys183Asp, Lys198Glu.
 CC The RENP, which selectively and specifically binds
 CC lipopolysaccharide (LPS) and has endotoxin neutralising activity,
 CC can be used to detect a site of Gram negative bacterial infection,
 CC and prevent or treat endotoxin related disorders, preferably where
 CC LPS mediated stimulation of neutrophils and mononuclear cells is
 CC inhibited, e.g. shock, disseminated intravascular coagulation,
 CC anaemia, thrombocytopenia, adult respiratory distress syndrome,
 CC renal failure, liver disease and conditions associated with Gram
 CC negative bacterial infection. The RENP has an enhanced serum half
 CC life relative to naturally occurring BPI. The RENP binds without
 CC triggering a significant, undesirable immune response.
 CC N.B. Sequence not given in specification, but produced using the
 CC wild type hbPI sequence given in figure 3.
 XX
 SQ Sequence 487 AA;
 Query Match 7.7%; Score 82; DB 17; Length 487;
 Best Local Similarity 20.7%; Pred. No. 5.5;
 Matches 48; Conservative 46; Mismatches 80; Indels 58; Gaps 11;
 QY 5 LILAAVAFVAVSADPIHYDKTEE-INKAIDDAIAEQSETIDPMKVPDHADKFE-RHY 62
 DB 20 VLVAIGTAVTAANVPCGVVRISSQGLDYASQGSTALQKE--LKRIKIDYSSFKIKHL 77
 QY 63 GIVDFKGLAMRNIEAR-----GLKQMKRQGANVKGEGCIYKAHLILGV 107

```

XX OS Chlamydia pneumoniae.
XX PN MO9927105-A2.
XX PD 03-JUN-1999.
XX PF 20-NOV-1998; 98WO-IB01890.
XX PR 04-NOV-1998; 98US-0107078.
XX PR 21-NOV-1997; 97FR-0014673.
XX PA (GENSET ) GENSET.
XX PI Griffiths R;
XX DR WPI. 1999-357842/30.
XX PS Genome sequence of Chlamydia pneumoniae
XX PS Page 1240-1241; Disclosure; 1912pp; English.
XX CC AA34584-Y35879 represent the proteins encoded by all the open reading
XX CC frames in the complete genome (see AA349190) of Chlamydia pneumoniae.
XX CC C. pneumoniae causes respiratory disease such as pneumonia and
XX CC bronchitis and is thought to be a contributing factor in heart
XX CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX CC nodosum or pharyngitis. The polypeptides encoded by the open reading
XX CC frames of the C. pneumoniae genome (see AA34584-Y35879) can be used in
XX CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX CC nucleotide sequences can also be used as immunogenic compositions,
XX CC especially where the vector directs the expression of a neutralising
XX CC epitope of C. pneumoniae.
XX SQ Sequence 397 AA;

Query Match 7.7%; Score 82; DB 20; Length 397;
Best Local Similarity 20.4%; Pred. No. 4.1;
Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;

QY 23 DKITEE---INKAIDDAIAEQSETIDPMKVPDHAKFE-----RH 61
DB 135 NKFTHOGIRILTKA---SISAIEESQNVRTVNDQVEEDVYLAIGQFWTASIGLDN 211
QY 62 VGIV-DFKGLAMRNIEAGLKQMKRGD-----ANKGEGGIYKAHLIGVHDDIYS 113
DB 212 AGVIDDDGVIPEDEMTNRNPNVIAIGDITGKWLAAHVAHQGYAAKNISGHHE---V 268
QY 114 MEYDLAYKLGDLHPPTTHVVISDIQDFVALSLSEISDEGNI--TMTSEVVRQ----- 161
DB 269 MDYSAIPSVYIFTHP-----EIAMVGLSLQEAEGQONLPKILTKRPFPAIGKAVAGAS 320
QY 162 --FANVVNH-----IGLSILDPFEGVLSLVLTALFQDTVRKEMT 199
DB 321 DGFAAIVSHETQOILGAVVIGPHASSLIGEMTL-----AIRNELT 361

RESULT 47
AAB62188
ID AAB62188 standard; Protein; 461 AA.
XX AC AAB62188;
XX DT 11-JUN-2001 (first entry)
XX DE C. pneumoniae lpda protein.
XX KM Chlamydia; lpda; infection; diagnostic; medicament; vaccine;
XX KW gene therapy; antibacterial.
XX OS Chlamydia pneumoniae.
XX PN WO200121802-A1.

```

```

XX PD 29-MAR-2001.
XX PF 15-SEP-2000; 2000WO-CA01086.
XX PR 17-SEP-1999; 99US-0154325.
XX PA (AVET ) AVENTIS PASTEUR LTD.
XX PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX DR WPI. 2001-257992/26.
XX DR N-PSDB; AAF57425.
XX PS Novel Chlamydia pneumoniae lpda protein and polynucleotides encoding
XX PS them useful as component of vaccines for treating Chlamydia infections,
XX PS and for detecting Chlamydia infection in the body fluid of a mammal.
XX PS Claim 16; Fig 1; 78pp; English.
XX CC This represents the Chlamydia pneumoniae lpda protein. The lpda protein
XX CC can be expressed by standard recombinant methodology. The lpda protein,
XX CC polynucleotide and antibodies specific to the protein and vaccines
XX CC comprising the lpda protein are useful for preventing or treating
XX CC Chlamydia (C. trachomatis, C. psittaci, C. pneumonia or C. pecorum)
XX CC infection. Vaccine vectors, and the lpda protein are useful in the
XX CC preparation of medicaments for preventing and/or treating Chlamydia
XX CC infection. Primers and probes derived from the lpda gene are also useful
XX CC for detecting and/or identifying Chlamydia in the biological material.
XX SQ Sequence 461 AA;

Query Match 7.7%; Score 82; DB 22; Length 461;
Best Local Similarity 20.4%; Pred. No. 5.1;
Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;

QY 23 DKITEE---INKAIDDAIAEQSETIDPMKVPDHAKFE-----RH 61
DB 219 NKFTHOGIRILTKA---SISAIEESQNVRTVNDQVEEDVYLAIGQFWTASIGLDN 275
QY 62 VGIV-DFKGLAMRNIEAGLKQMKRGD-----ANKGEGGIYKAHLIGVHDDIYS 113
DB 276 AGVIDDDGVIPEDEMTNRNPNVIAIGDITGKWLAAHVAHQGYAAKNISGHHE---V 332
QY 114 MEYDLAYKLGDLHPPTTHVVISDIQDFVALSLSEISDEGNI--TMTSEVVRQ----- 161
DB 333 MDYSAIPSVYIFTHP-----EIAMVGLSLQEAEGQONLPKILTKRPFPAIGKAVAGAS 384
QY 162 --FANVVNH-----IGLSILDPFEGVLSLVLTALFQDTVRKEMT 199
DB 385 DGFAAIVSHETQOILGAVVIGPHASSLIGEMTL-----AIRNELT 425

RESULT 48
ABB94271
ID ABB94271 standard; Protein; 461 AA.
XX AC ABB94271;
XX DT 05-JUN-2002 (first entry)
XX DE Chlamydia pneumoniae protein sequence SEQ ID NO:399.
XX KM Chlamydia; infection; Chlamydia; vaccine; detection; diagnosis;
XX KW antigen; antibacterial; immunostimulant; immune response;
XX OS Chlamydia pneumoniae.
XX PN WO200208267-A2.
XX PD 31-JAN-2002.

```

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
XX Staphylococcus epidermidis.
OS WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAXO) GLAXO GROUP LTD.
XX
XX kimmerly WJ;
XX WPI: 2001-316495/33.
XX N-PSDB; AAH53297.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 540; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 277 AA:
SQ
Query Match 7.7%; Score 82; DB 22; Length 277;
Best Local Similarity 24.4%; Pred. No. 2.5;
Matches 31; Conservative 27; Mismatches 39; Indels 30; Gaps 7;
OY 84 MKRGDANVKGEGIVKAKHLIGVHDIYSMEYDLAYKIGDLHP-TNVISDIQDFVAL 142
DB 131 VKRE-----KGORGL-----DILNQLYDIDYPTRVINPQSH--SDIDTLLIKL 172
OY 143 SLEISDEGNITMTSPFVQFANVNVHIGLSLIDPIFGVLSVYLTAIFODPYRKEMTKVL 202
DB 173 AQOY--HAHVITTDPNL-----NKVCHVOGITALN-----VNDLSAIPKPNVHQGDQLSTL 221
OY 203 APAFKRE 209
DB 222 LTKIGKE 228

RESULT 45
ABP39212
ID ABP39212 standard; Protein; 362 AA.
XX
XX AC ABP39212;
XX DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4057.
DE
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
XX Staphylococcus epidermidis.
OS US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI: 2002-381255/41.
XX N-PSDB; ABN91757.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure: SEQ ID 4057; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPRO web site.
XX
XX Sequence 362 AA:
SQ
Query Match 7.7%; Score 82; DB 23; Length 362;
Best Local Similarity 24.4%; Pred. No. 3.6;
Matches 31; Conservative 27; Mismatches 39; Indels 30; Gaps 7;
OY 84 MKRGDANVKGEGIVKAKHLIGVHDIYSMEYDLAYKIGDLHP-TNVISDIQDFVAL 142
DB 216 VKRE-----KGORGL-----DILNQLYDIDYPTRVINPQSH--SDIDTLLIKL 257
OY 143 SLEISDEGNITMTSPFVQFANVNVHIGLSLIDPIFGVLSVYLTAIFODPYRKEMTKVL 202
DB 258 AQOY--HAHVITTDPNL-----NKVCHVOGITALN-----VNDLSAIPKPNVHQGDQLSTL 306
OY 203 APAFKRE 209
DB 307 LTKIGKE 313

RESULT 46
AAV35476
ID AAV35476 standard; Protein; 397 AA.
XX
XX AAV35476;
XX
XX 13-SEP-1999 (first entry)
XX
XX Chlamydia pneumoniae transmembrane protein sequence.
DE
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.

OS Lawsonia intracellularis.
XX
XX WO9720050-A1.
XX
XX 05-JUN-1997.
XX
XX 29-NOV-1996; 96WO-A000767.
XX
XX 30-NOV-1995; 95AU-0006911.
XX
XX 30-NOV-1995; 95AU-0006910.
XX
XX (DARA-) DARATECH PTY LTD.
XX (PIGR-) PIG RES & DEV CORP.
XX
XX Hasse D, Panaccio M;
XX
XX WPI: 1997-310605/28.
XX
XX N-PSDB; AAT69201.
XX
XX Vaccine for treating or preventing Lawsonia intracellularis
XX infection - especially in pigs, containing non-pathogenic form of
XX bacterium or its components
XX
XX
XX Claim 10; Page 38-42; 94pp; English.
XX
XX The GroEL heat shock protein (AAW16678) of Lawsonia intracellularis
XX can be used in vaccines to protect birds and animals against
XX intestinal diseases, esp. to protect pigs against porcine
XX proliferative enteropathy. It is the expression product of a DNA
XX molecule (AAW69201) obtd. by screening an L. intracellularis library
XX with rabbit anti-L. intracellularis sera. GroEL, GroES (AAW16679)
XX and other L. intracellularis polypeptides (AAW16680-85) can be
XX administered as recombinant polypeptides or expressed as
XX recombinant vaccines utilising bacterial, fungal or viral vectors.
XX Antibodies raised against the polypeptides may be useful in
XX immunotherapy, diagnosis of infection and detection.
XX
XX Sequence 548 AA;
SQ
Query Match 7.7%; Score 82.5; DB 18; Length 548;
Best Local Similarity 18.7%; Pred. No. 5.7;
Matches 49; Conservative 43; Mismatches 83; Indels 87; Gaps 10;
QY 10 VAFVAVSADPIHYDKTEINKAIDDAIAIEQSEITDPMKVPDHADFEERHVGIVDFKG 69
DB 104 VKLVAAGRNP-----AIKRGIDKAVVAV--TKELSDITKPTDQKEIAQVGTISANS 154
QY 70 ELAMRNITARGIKOMKRGODANVKEGEGIVKAHLIGVHDIIVSMEDLAY----- 120
DB 155 DTIGNITIAEMAKVKGKGVITVEAKG-----LETTLDVVEGKMPRGYISPYFVTNP 208
QY 121 -----KLGDLPPTTHVVISDIQDFVALSLDISDEGNTTMTSFEVR 160
DB 209 EKAVCELDNPTLCNEKRTITSMKMDPLIEQYAKVNRPLIIIAEVEGBALATL----- 262
QY 161 QFANYVNHIG-----LSILDPIFG-----VLSD--VLT--AIFOD----- 192
DB 263 ----VVNKLRGALGVAVKAPGFGERRKAMLEDIILITGGEALIEFDGRGKLENVSLSLIG 318
QY 193 -----TVRKEMTKVLAPAFKRE 209
DB 319 TAKRVVIDKENTTIVDGAGKSE 340

RESULT 43
AAU32848
ID AAU32848 standard; Protein: 2486 AA.
XX
XX AAU32848;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human secreted protein #3339.
DE

XX
XX Human; vaccination; gene therapy; nutritional supplement;
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
XX Homo sapiens.
XX
XX WO200179449-A2.
XX
XX 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US08656.
XX
XX 18-APR-2000; 2000US-0552929.
XX
XX 26-JAN-2001; 2001US-0770160.
XX
XX (HYSE-) HYSEQ INC.
XX
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-611725/70.
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy -
XX
XX
XX Claim 20; Page 678; 765pp; English.
XX
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
XX with altered levels of polypeptide. The polypeptides are also useful for
XX identifying agents (agonists and antagonists) that bind to them. Cells
XX expressing the proteins are useful for identifying a therapeutic agent
XX for use in treatment of a pathology related to aberrant expression or
XX physiological interactions of the polypeptide. Vectors comprising
XX the nucleic acids encoding the polypeptides and cells genetically
XX engineered to express them are also useful for producing the proteins.
XX The proteins are useful in genetic vaccination, testing and
XX therapy, and can be used as nutritional supplements. They may be used to
XX increase stem cell proliferation; to regulate haematopoiesis; and in
XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX immune suppression and/or stimulation; as anti-inflammatory agents; and
XX in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
XX sequences of novel human secreted proteins of the invention.
XX
XX Sequence 2486 AA;
SQ
Query Match 7.7%; Score 82.5; DB 22; Length 2486;
Best Local Similarity 24.2%; Pred. No. 49;
Matches 39; Conservative 32; Mismatches 49; Indels 41; Gaps 7;
QY 45 TIDPKKVPDHDHDKFERHVGIVDFKGL-----AMRIEARGLKOKKRG----- 88
DB 2238 TINPLCIENYADK-ESRGVLEPESTVEIKFRKKDLIKSMKRIDPAYKKLWQLEPDL 2296
QY 89 DANVGESEIVKAHLIGVHDIIVSMEDLAYKLGLDLPHT-----TVVISDIDQFVAL 142
DB 2297 DKDRKDLGRKLAR-----EDLLRPIYHOVANGFADPHDPTFRMLEKEVSDILEMKRAR 2351
QY 143 S-----LEISDEGNTTMSFEVRQFANYVNHIGLSIL 175
DB 2352 TFLYWRRLRLLEDO-----VKQETLOASGELSHVHQSML 2387

RESULT 44
AAG82447
ID AAG82447 standard; Protein: 277 AA.
XX
XX AAG82447;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO:1988.
DE

ID AAU71861 standard; Protein; 845 AA.
 XX
 AC AAU71861;
 XX
 DT 26-FEB-2002 (first entry)
 XX
 DE Leishmania antigen 4G2-83 extended protein.
 XX
 KW Leishmaniasis; Leishmania antigen; immunostimulant; protozoa; Ldp23;
 KM Interleukin-15; Lbhsp83; M15; Lt-1; LbelFAA; LmSP1a; LmSP9a; MAPS-1A;
 KM LmSP1; LmSP8; LmSP5; LmSP9; LmSP13; LmSP19; LcGSP1; LcGSP3;
 KM LcGSP4; LcGSP8; LcGSP10; 1G6-34; 1E6-44; 4A5-63; 1B11-39; 2A10-37;
 KM 4G2-83; 4H6-41; 8G3-100.
 XX
 OS Leishmania major.
 XX
 PN WO200179276-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 05-APR-2001; 2001WO-US11254.
 XX
 PR 14-APR-2000; 2000US-0551974.
 PR 05-MAY-2000; 2000US-0555501.
 PR 14-AUG-2000; 2000US-0639206.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Campos-Neto A, Webb JR, Dillon DC, Skeiky YAM, Bhatia A;
 PI Coler RM, Probst P;
 XX
 DR WPI; 2002-061971/08.
 DR N-PSDB; AAS96071.
 XX
 PT New isolated Leishmania antigens, useful for prevention, treatment and
 PT diagnosis of leishmaniasis, also related nucleic acids for genetic
 PT vaccination
 XX
 PS Claim 1; Page 190-192; 193pp; English.
 XX
 CC The invention relates to polypeptides comprising an immunogenic part of a
 CC Leishmania antigen. The Leishmania polypeptides and their associated DNA
 CC sequences, epitopes and fusion proteins are used in the production of
 CC compositions used for inducing a protective immune response against
 CC leishmaniasis, for prevention and treatment of the disease. The
 CC compositions can also be used generally to treat diseases that respond to
 CC interleukin-15 stimulation. In addition, the products may contain an
 CC immunostimulant. Sequences AAU71805-AAU71862 represent Leishmania
 CC antigens and antigenic peptides of the invention.
 CC
 XX
 SQ Sequence 845 AA;
 Query Match 7.8%; Score 83; DB 23; Length 845;
 Best Local Similarity 23.3%; Pred. No. 9.4;
 Matches 42; Conservative 34; Mismatches 70; Indels 34; Gaps 10;
 OY 29 INKAIDDAIAAEQSETPMKVPRHADKFERHVGIVDFKGLAMKNIEARGIKOKKROG 88
 DB 2 VNFYDVQVRELMDYRDPQIRKMSVIAHVD-----HCKSTLSDSLGAAGIKKEEG 52
 OY 89 DANV---KGEIGIVKAHLIGVHDDIVSMEDLAYK-LGDLHPPTHVTSIDQFVVALSL 144
 DB 53 DKRINDTRADE-IARG---TITKSTALSMHYHVRKEMIDG-----DDKRDPLNLN-- 100
 OY 145 EISDGNITMTSFEVROFANVNHIGLSILDFIGVLSVDVLTALFQDVRKEMKVLAP 204
 DB 101 -IDSPGHVDFSS-EVTAALRLYTD--GALVVYDCVGVGVQVETETVL-----ROALTEIRIP 151

AC ABB53805;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein dnaE.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 KM
 OS Lactococcus lactis IL1403.
 XX
 PN FR2807446-A1.
 PD 12-OCT-2001.
 PF 11-APR-2000; 2000FR-0004630.
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 XX
 DR WPI; 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species -
 XX
 PS Claim 6; SEQ ID No 507; 2504pp; French.
 XX
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB53621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 1060 AA;
 Query Match 7.8%; Score 83; DB 23; Length 1060;
 Best Local Similarity 21.2%; Pred. No. 13;
 Matches 32; Conservative 29; Mismatches 38; Indels 52; Gaps 8;
 OY 91 NVKGE---EGIVKAHLIGVHDDIVSMEDLAYK-----LGDLHPPTHVTSIDQF-- 138
 DB 6 NTKTEYSFLDSYVKV-----DDYLEFAHRLGYOTVGICDGNLHAFFVRKAOKFNL 58
 OY 139 ---VVALSLEISDEG-----NT---TMTSFEVROFANVNHIGLSILD 176
 DB 59 OPTISIELFEMRGRLPIAFSTAKDTBGYKNLRLSTLHNVRROFSDIQNHLGIALTI 118
 OY 177 P-IFGVLSVDVLTALFQDVRKEMKVLAPAF 206
 DB 119 PETYGSLSL-----TELSVADAEAF 139

RESULT 41
 ABB53805
 ID ABB53805 standard; Protein; 1060 AA.
 XX

RESULT 42
 AAM16678
 ID AAM16678 standard; Protein; 548 AA.
 XX
 AC AAM16678;
 XX
 DT 20-AUG-1997 (first entry)
 XX
 DE Lawsonia intracellularis GroEL.
 KM Intestinal disease; porcine proliferative enteropathy; vaccine;
 KM GroEL; Immunotherapy; antibody; diagnosis; heat shock protein.
 XX

Db 402 A 402

RESULT 38

ABBS3885
ID ABBS3885 standard; Protein: 775 AA.XX
AC ABBS3885;XX
DT 16-MAY-2002 (first entry)XX
DE Lactococcus lactis protein yfGQ.XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.XX
OS Lactococcus lactis IL1403.XX
PN FR2807446-A1.XX
PD 12-OCT-2001.XX
PE 11-APR-2000; 2000FR-0004630.XX
PR 11-APR-2000; 2000FR-0004630.XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.XX
PI Boloiline A, Sorokine A, Renault P, Ehrlich SD;XX
DR WPI; 2002-043418/06.XX
PT New nucleotide sequence useful in the identification or Lactococcus
PT lactis and related species -XX
PS Claim 6; SEQ ID No 587; 2504bp; French.

XX The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABBS3885) and related proteins (ABBS3300-ABBS5621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 775 AA;

Query Match 7.9%; Score 84; DB 23; Length 775;

Best Local Similarity 19.8%; Pred. No. 6.5; Mismatches 103; Indels 62; Gaps 6;

Matches 50; Conservative 37;

XX 22 YDKIEEINKAIDDAIAIEOSEITDPMKVPDADKFERHVGIVDFKGL-----AMRN 75

XX 397 YEKVSEINHLLEEGFRVYLAGTKE--KIYQNLNGAVALGYVLIANIRENAKSTFNY 454

XX 76 IEARGLKQKKROG-----ANVKGEGIVKAHLL----- 104

XX 455 FAEOGVNIKIVISGDMPOFSAVAKRAGITGAERFIDANILKTKEDIDQAVESYVFGRYT 514

XX 105 -----IGVHDDIVSMEDY-----LAYKLGDLHPTTHYISDIODFVVALSLSEISDE 149

XX 515 PQOKRLVQALRKRDHYAMTGDVNDILAMKSADCSIMASGDADANTQAVVLDSDP 574

XX 150 GNTITSFEVROFANVNVHIGLSILDPIFGVLSDVTAIFODTYARKEMTK-----V 201

XX 575 GIMTQVVEGRVNVNVQGSATLFLVKNLFTIILAIISIFFTYPLQASQSLISLFTI 634

XX 202 LAPAFKRELEKN 213

XX 635 GIPGFLLSLEEN 646

RESULT 39

AAG89772
ID AAG89772 standard; Protein: 412 AA.XX
AC AAG89772;XX
DT 26-SEP-2001 (first entry)XX
DE C glutamicum protein fragment SEQ ID NO: 3526.XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.XX
OS Corynebacterium glutamicum.XX
PN EP1108790-A2.XX
PD 20-JUN-2001.XX
PE 18-DEC-2000; 2000EP-0127688.XX
PR 16-DEC-1999; 99JP-0377484.XX
PR 07-APR-2000; 2000JP-0159162.XX
PR 03-AUG-2000; 2000JP-0280988.XX
PA (KIOW) KIOWA HAKKO KOKYO KK.XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;XX
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;XX
DR WPI; 2001-376931/40.XX
DR N-PSDB; AAH64991.XX
PS Claim 17; SEQ ID NO: 3526; 246bp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX

SQ Sequence 412 AA;

Query Match 7.8%; Score 83.5; DB 22; Length 412;

Best Local Similarity 31.5%; Pred. No. 3; Mismatches 41; Indels 17; Gaps 6;

Matches 34; Conservative 16;

XX 54 HADKFERHNGIYDFGELAMRNIEARGIKQMKROGDANVKGEGIVKAHLLIGVHDDIYS 113

XX 134 YAEAFD-----IOPFSGEDSEFTVLKSGDFEDALDQ-----QWEGSNRPYAAAGVSDN-VS 183

XX 114 MEYDLAYKLGDLHPTTHVISDI-----QDFVALSLSEISDEGNITMTS 156

XX 184 YTYIDNRPICD-RVTSVTITDDIPDPDERNDYVVAASLYL-QSGNDEGMA 229

RESULT 40

AAU71861

```

XX 02-JUL-2002 (first entry)
DT Streptococcus polypeptide SEQ ID NO 3268.
XX
DE Streptococcus polypeptide SEQ ID NO 3268.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX Streptococcus agalactiae.
OS
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI: 2002-352536/38.
XX N-PSDB: ABN67677.
XX
XX New streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX
XX Claim 1: Page 3480-3481; 4525pp: English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX
XX Sequence 306 AA:
SQ

```

Query Match 7.9%; Score 84; DB 23; Length 306;
Best Local Similarity 23.8%; Pred. No. 1.7;
Matches 34; Conservative 25; Mismatches 62; Indels 22; Gaps 6;

```

OY 18 DPHYDKITEELINKAI-DAIAIAESETIDPMKVPDADKFERHVGI-VDFKGLAMRN 75
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 146 DIFFEMLCETLEKNSIRQEQYTLQOEGLDTRKIKELSNYFANQIYISDFVGNL--EN 203
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 76 IEARGIAKQKROGDANVKEGIVKAHLIGVDDIYSMEYDLAYKLG-----DLHPT 128
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 204 LSKDKIKQEKVE-----SYERLKDSPQAG-----VEVSYDSNYLGLFLINTLADSQTT 252
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 129 THVYSDIOEVALSLEISDEGN 151
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 253 RQLSDMSQATVGMFPFSQFHEGH 275
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

RESULT 37
AAVS9996
ID AAVS9996 standard: Protein; 410 AA.
XX
XX AAVS9996;
XX
XX 31-JAN-2000 (first entry)
DT
XX
XX Human endometrium tumour EST encoded protein 56.
DE
XX
XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:
KW treatment; uterine; gene therapy; expressed sequence tag.
XX
XX Homo sapiens.
OS
XX
XX DE19817948-A1.
XX
XX 21-OCT-1999.
XX
XX 17-APR-1998; 98DE-1017948.
XX
XX 17-APR-1998; 98DE-1017948.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
XX WPI: 1999-591957/51.
XX N-PSDB: AAZ41999.
XX
XX New nucleic acid sequences expressed in uterine cancer tissues, and
XX derived polypeptides, for treatment of uterine and endometrial cancer
XX and identification of therapeutic agents -
XX
XX Claim 23: Page 297; 444pp: German.
XX
XX This invention describes novel human nucleic acid (cDNA) sequences (A),
XX that are highly expressed in uterine tumour tissue and which have
XX anticancer and cytostatic activity. (A) are used (i) for recombinant
XX expression of polypeptides (B) and (ii) to isolate complete genes. (B)
XX are used (i) to identify agents suitable for treatment of uterine or
XX endometrial cancer; (ii) directly for treating these forms of cancer
XX (including expression from gene therapy vectors) and (iii) for
XX generation of specific antibodies. (A) are identified by assembling ESTs
XX (expressed sequence tags) from a particular tissue type before comparison
XX of expression patterns. This allows a significantly longer fragment of
XX the gene to be revealed, so should reduce the number of failures
XX associated with the fact that ESTs from different libraries may represent
XX different parts of the same unknown gene, distorting the estimated
XX frequency of occurrence in a particular tissue. AAVS9941-Y60328 represent
XX protein fragments encoded by the human endometrium tumour cDNA library
XX derived EST fragments represented in AAZ41981-242121.
XX
XX Sequence 410 AA:
SQ

```

Query Match 7.9%; Score 84; DB 20; Length 410;
Best Local Similarity 22.1%; Pred. No. 2.6;
Matches 40; Conservative 33; Mismatches 62; Indels 46; Gaps 9;

```

OY 15 VSADPHYDKITEELINKAIDAIAIESETIDPMKVPDADK-KEERHVGIV--DFKGE 70
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 246 VSAEKYK--TSHVNGITEADPTIYSGKIVIPRLSPVPTQREYOGMIEIVEGDMKGE 302
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 71 LAMRNIEARGIAKQKROGDANVKEGIVKAHLIGVDDIYSMEYDLA----- 119
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 303 -----VYPPGIVGMANKGDLQKES--VAFQLCV-LGQNAQMAIANIPLPRATYECVK 354
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 120 -----YKLGDLPTTHVYSDIOEVALSLEISDEGNITTSFEVROFANVNHIGL 172
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 355 DQGFINTYEVDGSKLFFHKEVOD---GIELQAGE-----VERSVLPKSSGGL 401
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 173 S 173
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

XX		Protein involved in intermediate metabolism of polypeptides.	
DE			
XX		Vaccine; eye disease; conventional trachoma; nonendemic trachoma;	
KW		paratrachoma; inclusion conjunctivitis; genital disease; perithenarthritis;	
XK		nongonococcal urethritis; epididymitis; cervicitis; salpingitis;	
KM		bartholinittis; pneumopathy; venereal lymphogranulomatosis.	
OS		Chlamydia trachomatis.	
XX			
PN		MO928475-A2.	
PD		10-JUN-1999.	
XX			
PF		27-NOV-1998; 98WO-IB01939.	
XX			
PR		04-NOV-1998; 98US-0107077.	
PR		28-NOV-1997; 97ER-0015041.	
PR		17-DEC-1997; 97ER-0016034.	
XX		(GEST) GENSET.	
PA			
XX		Grieffals R;	
PI			
XX		WPI: 1999-371125/31.	
PS			
XX		Disclosure: Page 739-741; 1755pp; English.	
CC			
CC		AA36754-Y37949 are encoded by open reading frames (ORFs) of the genome	
CC		of Chlamydia trachomatis (see AA201425). The polypeptides can be used as	
CC		vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences	
CC		can also be used to control growth of the microorganism. Chlamydia	
CC		trachomatis is responsible for a large number of diseases, e.g. eye	
CC		diseases such as conventional trachoma, nonendemic trachoma,	
CC		paratrachoma, and inclusion conjunctivitis; genital diseases such as	
CC		nongonococcal urethritis, epididymitis, cervicitis, salpingitis,	
CC		perithenarthritis, bartholinittis; pneumopathy in breast feeding infants;	
CC		and venereal lymphogranulomatosis. The polypeptides of the invention	
CC		may be of use in treating these diseases.	
XX			
SQ		Sequence 1396 AA:	
		Query Match 7.9%; Score 84.5; DB 20; Length 1396;	
		Best Local Similarity 23.6%; Pred. No. 13;	
		Matches 49; Conservative 34; Mismatches 66; Indels 59; Gaps 12.	
OY		43 SETIDPKVPVDHADKFERRHGVIVDFKGELAMRNIT-----EARGLKQ-----MKRQ 87	
Dd		: : : : : : : : : : : : : :	
Db		1145 AELVEARPEEDPAD--IAKIDGVDEFKGIQKNRKRLVVRDELGTGMEEHLISLTKHLIYQR 1203	
OY		88 GDANKGE---EGYIKAKHLII---GVHD-DIVSMEDYLAVKLGDLHPPTHYSIDIQFV 139	
Dd		:: :::: : : : : : : : : : :	
Db		1204 GDSVIKGOQLDGLVAIPHILAIIGCVRELKVLVNEVGVEYRLQGV-----DINCKH 1255	
OY		140 VAL-----SLEISDEGNITM-----TSFEIROFAVV---VNHIIG-----LSIDP 177	
Dd		: : : : : : : : : : : : : : :	
Db		1256 IEIIVRQMLOKRVRIIDPDPTILLBESEDVKKKFYDENNRTEEDGGKPRAOVAVPLGITKA 1315	
OY		178 IFGYLDVLTALAFODTVRRKEMTKLAPA 205	
Dd		: : : : : : : : : :	
Db		1316 SLGTESFISAASFQDT-----TRVLTLDA 1338	
RESULT 35			
ID	ABB58245	standard; Protein: 1887 AA.	
AC	ABB58245:		
XT	26-MAR-2002	(first entry)	
XX			

```

DE Drosophila melanogaster polypeptide SEQ ID NO 1527.
XX
KW Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL02348.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 1527; 21np + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB12072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX
Sequence 1887 AA:
XX
Query Match 7.9%; Score 84.5; DB 22; Length 1887;
Best Local Similarity 22.7%; Pred.No.20;
Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12.
QY 17 ADPIHYDKTEINKAIDDAIAIQTSETIDPMKYVDNA--DKFERHGVY----DFRG 69
DB 666 ADPQTYNEIQQAIRKAKDDVINVIQKAHMELEPPGNTLRQTFENKVRRIINDARDKDG 745
QY 70 ELAMNI-EARLKKMKMRQ-----GDANVKGEE--GIVKAHLIGVHD 109
DB 746 GSAKSLTEYNNLKAQVSVSGSKGSININISQVIACVGGQOVESKRIPIYGRKRTLPDFTKD 805
QY 110 D-----IVSMEDYDLAKVIGDLAPTT--AVISIDQDFVVALSLDISDEGNI-----T 153
DB 806 DYGPESRGFEVMSY-----LAGLTPSELYFHAMCG-REGIDITAVKTAETGIIQRLIKA 859
QY 154 MTSFEVROFANVNVHIGGLSTL---DPIFGVLSD----- 184
DB 860 MESVMVNVNDGTVRNSVGLQILRYGEDGICGLAEVFOQMPVTKLNSKSPFKRFKFDWSNE 919
QY 185 -VLTALFODTVAKEMT 199
DB 920 RLMKRVFTDDVIKEMT 935
RESULT 36
ABP27046
ID ABP27046 standard; Protein; 306 AA.
XX
XX
XX ABP27046;

```


KW vaccination: endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimerly MJ;
XX
DR WPI: 2001-316495/33.
XX
DR N-PSDB: AAH53067.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX
PS Claim 18; Page 432; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 817 AA;
XX
Query Match 7.9%; Score 84.5; DB 22; Length 817;
Best Local Similarity 23.0%; Pred. No. 6.2;
Matches 50; Conservative 35; Mismatches 87; Indels 45; Gaps 11;
QY 18 DP:-----YDKITEINKAIDAIAIEQSEITIDPMKVPDHDKFERHVGIVDFKGEA 72
DB 504 DT:HKRVYIGONDVANSISKAVRARAGLK-----DP-----KRLGSGTIFLGPTG 548
QY 73 MRNIE-ARGL-KOMKQGDANVKE--EGIVKAHL---LIGV-----HDDIVSMEDLAY 120
DB 549 VGTTELARALAESMFEGDDAMIRVDSEFEKKAHVSRLVGAAPGYGHDDGGLTEKVR 608
QY 121 KLGDLHPHTHVISD-----IODEFVALSLEISDGNITMTSFEVROPANVY----NHIGL 172
DB 609 K-----PYSVILDEIEKAPDVFNILQVLDGHLDTGRVDFRNTVIIMTSNVAQ 663
QY 173 SILDPIFGVLSVLTALFODTVRKEMTKVLAFAFKRE 209
DB 664 ELQDQRFAGFGGASGSDYETVVKTKMKELKNSFRPE 700

RESULT 33
ABP39236
ID ABP39236 standard; Protein; 823 AA.
XX
AC ABP39236;

XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4081.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
XX
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI: 2002-381255/41.
XX
DR N-PSDB: ABN91781.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
PS Disclosure; SEQ ID 4081; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 823 AA;
XX
Query Match 7.9%; Score 84.5; DB 23; Length 823;
Best Local Similarity 23.0%; Pred. No. 6.2;
Matches 50; Conservative 35; Mismatches 87; Indels 45; Gaps 11;
QY 18 DP:-----YDKITEINKAIDAIAIEQSEITIDPMKVPDHDKFERHVGIVDFKGEA 72
DB 510 DT:HKRVYIGONDVANSISKAVRARAGLK-----DP-----KRLGSGTIFLGPTG 554
QY 73 MRNIE-ARGL-KOMKQGDANVKE--EGIVKAHL---LIGV-----HDDIVSMEDLAY 120
DB 555 VGTTELARALAESMFEGDDAMIRVDSEFEKKAHVSRLVGAAPGYGHDDGGLTEKVR 614
QY 121 KLGDLHPHTHVISD-----IODEFVALSLEISDGNITMTSFEVROPANVY----NHIGL 172
DB 615 K-----PYSVILDEIEKAPDVFNILQVLDGHLDTGRVDFRNTVIIMTSNVAQ 669
QY 173 SILDPIFGVLSVLTALFODTVRKEMTKVLAFAFKRE 209
DB 670 ELQDQRFAGFGGASGSDYETVVKTKMKELKNSFRPE 706

RESULT 34
AAY36871
ID AAY36871 standard; Protein; 1396 AA.
XX
AC AAY36871;
XX
DT 07-OCT-1999 (first entry)

XX New nucleotide sequence useful in the identification or Lactococcus
PT lactic and related species -
XX
XX
PS Claim 6, SEQ ID No 1407; 2504pp: French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 878 AA;

Query Match 8.0%; Score 85; DB 23; Length 878;
Best Local Similarity 24.7%; Pred. No. 6;
Matches 39; Conservative 33; Mismatches 44; Indels 42; Gaps 8;

QY 72 AMRNIARGLKQKRGDANVKGEGIVKA-HLLIGVHDDIVSMEXDLAYKLDLHPPTH 130
DB 106 ALRQMSAPSAKYLK-----NCKTSIPARELVG---DIVSL-----AGDFIPADG 149
QY 131 VLSIDDFVALSLSEISDEGNITMTSFEVROFANV-----NHIGLSILDPHFG 180
DB 150 RLIDVONLNR-----EEGMLTGESEPEKFSVDIEGVALDRKNMVFSSSLV--VYG 200
QY 181 VLSVDLTAIFODTVRKEMTKVLA-----FKRELEK 212
DB 201 RADFLVTALAEQTEIGKIAQMLETAFAKQPPLOQKLEK 238

RESULT 29
AAR35439
ID AAR35439 standard; Protein; 293 AA.
XX
AC AAR35439;
XX
DT 23-AUG-1993 (first entry)
XX
DE B. burgdorferi strain Ip90 OsPB.
XX
KW OsPB; coding region; B. burgdorferi; strain; ACAI; North America;
KW tick; vector; Ixodes dammini; primer; B31; Ip90; class; I; II;
KW III; detection.
XX
OS Borrelia burgdorferi.
XX
FH Key Location/Qualifiers
FT Peptide 28..43
FT /note= "Immunogenic peptide, claim 42"
FT Peptide 61..73
FT /note= "Immunogenic peptide, claim 42"
FT Peptide 102..115
FT /note= "Immunogenic peptide, claim 42"
FT Peptide 232..244
FT /note= "Immunogenic peptide, claim 42"
FT Peptide 284..294
FT /note= "Immunogenic peptide, claim 42"
XX
XX WO9308306-A.
XX
XX PD 29-APR-1993.
XX
XX PF 22-OCT-1992; 92WO-US08972.
XX
XX PR 22-OCT-1991; 91US-0779185.
XX

PA (SYMB-) SYMBICOM AB.
XX
XX Barbour AG, Bergstrom S, Hansson L;
XX
XX
DR WPI: 1993-152497/18.
DR N-PSDB; AAO40697.
XX
XX
PT Detecting Borrelia burgdorferi by PCR-DNA analysis - using
PT defined nucleotide primer sequences, also a vaccine against Lyme
XX disease
XX
PS Disclosure; Fig 3b; 111pp; English.
XX
XX This sequence is encoded by the OsPB coding region of the B.
XX burgdorferi strain Ip90. The strain Ip90 is a Soviet isolate
XX from the tick vector, Ixodes persulcatus. The polypeptides
XX indicated correspond to regions within the B. burgdorferi OsPB
XX protein which are similar between strains B31, ACAI and Ip90
XX (see also AAR35437-39). These strains of B. burgdorferi are
XX representatives of the three classes of B. burgdorferi species,
XX class I, II and III. The immunogenic peptides may also be used in
XX vaccines to protect against Lyme disease. This sequence is
XX represented as it is given in the specification. It was not derived
XX from the corresponding DNA sequence.
XX
SQ Sequence 293 AA;

Query Match 7.9%; Score 84.5; DB 14; Length 293;
Best Local Similarity 22.3%; Pred. No. 1.4; Indels 51; Gaps 12;
Matches 52; Conservative 43; Mismatches 87; Indels 51; Gaps 12;

QY 1 MKRFL-----IAVAFVAVSADPIHYDKITEIRNK-----AIDAIAMAEQS 43
DB 1 MKRYLGFALVALLINACGCGAPKHNDDVEDLKKQKSKDPLVYEDVVKLFNN 60
QY 44 ETTDPKPKVPDADKFERHNGIYDFKGLAMRNITDANKLQKMGDANVKGEGIVKAHL 103
DB 61 E-IFISKEDKEDKYEIR-SYD-----KVELKGLSE-KNTGAGELEGKA-DKSKV 108
QY 104 LIGVHDI--VSME-YD-----LAVKLGDLHPPTHVSIQDPFVALSLSEISDEGN 151
DB 109 TMLVSDDLNTTLETYDPSKKTSSOVAKKQSLTETTYTSKLS-----AKKTRSN 162
QY 152 ITMTSFEVROFANVNH-----GSLSLDPHFGVLSVDVLAIFODTVRKEMTK 200
DB 163 TTLETYEMTDADNASKAVETLKNKGTLESGLVGKTTLTKECTVTLKKKLEK 215

RESULT 30
AAU58103
ID AAU58103 standard; Protein; 436 AA.
XX
AC AAU58103;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #18999.
XX
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX
OS Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX PD 01-NOV-2001.
XX
XX PF 20-APR-2001; 2001WO-US12865.
XX
XX PR 21-APR-2000; 2000US-199047P.
XX
XX PR 02-JUN-2000; 2000US-208841P.
XX

[illegible]

	PR	03-AUG-2000; 2000US-222898P.	
	XX	(UTAH) UNIV UTAH RES FOUND.	
	PA		
	PI	Placek L, White S, Fu Y, Skradski S;	
	XX	WPI: 2001-5889903/66.	
	DR	N-PSDB; AAD18286.	
	PT	Monogenic Audiogenic Seizure-Susceptible-1 (massl) genes, useful for	
	PP	producing animal models of audiogenic seizures -	
	PS	Claim 19; Fig 6; 79pp; English.	
	XX	The present invention relates to nucleic acid molecules designated	
	CC	monogenic audiogenic seizure-susceptible-1 (massl) genes. The nucleic	
	CC	acid molecule may be used via recombinant DNA methodologies in the	
	CC	production of transgenic animal (especially mouse) models for studying	
	CC	genetic abnormalities related to massl which result in seizure	
	CC	susceptible phenotypes (massl is audiogenic seizures). The present	
	CC	sequence is mouse monogenic audiogenic seizure-susceptible (massl)	
	CC	protein.	
	XX		
	SQ	Sequence 2780 AA;	
	Query Match	8.0%; Score 85.5; DB 22; Length 2780;	
	Best Local Similarity	21.4%; Pred. No. 28;	
	Matches	41; Conservative 32; Mismatches 78; Indels 41; Gaps 8	
OY	26	TTEINKAIDDAIAIEQSETDIPMKVPHHADKFERHV-----GIVDFEKLAMRNTGA 78 : :: : : : : : : : : : : : : : : Db 2508 TTVTAVILINDVNNAVIGVSQTASRSVIHGHEGMLOFHHVTTPRGNGNVMKVGVGGONLEV 2567	
OY	79	RGLQMRRKGOGANVGGE---EGIVKAHLLIGHDDIVSME-----YDLATYRKLGDLH 126 : :: :: :: :: :: :: :: :: : Db 2568 -----NFANFTFGQLFEFSBGTLNKTIFYVLHDNDIPEEKVEYQVVLYDV--KRQGS 2616	
OY	127	PTHVISDIODIFVALSLSEISDEG----NIWTLS-FEVROFANNV-----NHIGOLSTL 175 : : : : : : : : : : : : : : : : Db 2617 PAGVALLDAAQGVAALVTYEASDEDPHGVLNFLASSRFVYLQEANVTITLFVRNEFGSLGAI 2676	
OY	176	DPIFGVSDVLT 187 : : : : : Db 2677 NVTYATVPGIYS 2688	
	RESULT 28		
	ABB54705		
ID	ABB54705 standard; Protein; 878 AA.		
XX	ABBS4705;		
DT	16-MAY-2002 (first entry)		
DE	Lactococcus lactis protein yoab.		
KM	Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.		
OS	Lactococcus lactis IL1403.		
XX	FR2807446-Al.		
PD	12-OCT-2001.		
PF	11-APR-2000; 2000FR-0004630.		
PR	11-APR-2000; 2000FR-0004630.		
PA	(INRG) INRA INST NAT RECH AGRONOMIQUE.		
FI	Biotoline A, Sorokine A, Renault P, Ehrlich SD;		
RP	WPI: 2002-043418/06.		

CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 312 AA:

Query Match 8.1%; Score 86; DB 22; Length 312;

Best Local Similarity 25.0%; Pred. No. 1.1;
 Matches 32; Conservative 27; Mismatches 39; Indels 30; Gaps 7;

OY 84 MKRQGDANVKEGEGYKAHLIGVHDDIVSMEDYLAAYKGLDHP--THVYISDIQDFVVAL 142

DB 212 VKRE-----KGORGL-----DIINQLYDLDPYPRVHPPTQSH--SDIDTLILIKL 253

OY 143 SLEISDEGNITMSEFROFANVNVNHIGSLIDPIFGVLSVDLTAIFQDTYKREMTKYL 202

DB 254 AAOY--HAHVITTDENL-----NKVCHVGCTALN-----VNDLSEAIKPNVHOGDLSITL 302

OY 203 APAFKREL 210

DB 303 LTRICKEL 310

DE

XX

AAU75885

AC AAU75885;

DT 08-MAY-2002 (first entry)

XX Human adhesion molecule protein AD3/BAA25490.1.

XX Human; adhesion molecule; AD3; BAA25490.1;
 KW cardiovascular disease; atherosclerosis; ischaemia; thrombosis;
 KW haematological disease; leukaemia; blood clotting disorder; cancer;
 KW brain tumour; inflammatory disease; rhinitis; gastrointestinal disease;
 KW Crohn's disease; respiratory disease; asthma; immune disorder;
 KW rheumatoid arthritis; allergy; liver disease; cirrhosis; burn;
 KW endocrine disease; diabetes; bone disease; osteoporosis; wound healing;
 KW neurological disease; multiple sclerosis; bacterial infection;
 KW Mycobacterium tuberculosis infection; viral infection.

OS Homo sapiens.

XX WO200208423-A2.

PN 31-JAN-2002.

XX 24-JUL-2001; 2001WO-GB03318.

XX 24-JUL-2000; 2000GB-0018126.

PR 17-OCT-2000; 2000GB-0025447.

XX (INPH-) INPHARMATICA LTD.

PI Phelps CB, Fagan RJ, Gutteridge A;

DR WPI; 2002-155219/20.

XX N-PSDB; ABK14973.

PT Six novel proteins, termed KIAA0301, G7c, BAA25490.1, AAD21820.1,
 CAB01991.1 and CA17374.1, which have been identified as adhesion
 PT molecules, useful in the treatment and diagnosis of disease such as a
 XX cardiovascular disease, cancer and immune disorders -

PS Claim 1; Fig 25; 302pp; English.

XX The invention relates to six novel proteins, termed KIAA0301, G7c,
 CC BAA25490.1, AAD21820.1, CAB01991.1 and CA17374.1, which have been
 CC identified as adhesion molecules (also termed AD1-6), their
 CC encoding nucleic acids, fragments and functional equivalents.
 CC Also included are a vector comprising the nucleic acids, a host
 CC cell transformed with the vector, an AD-inhibitory ligand, a compound
 CC that either increases or decreases the level of expression or activity
 CC of the AD protein, a vaccine comprising the protein or nucleic acid,
 CC a method for the identification of a compound that is effective in
 CC the treatment and/or diagnosis of disease, comprising contacting AD
 CC or its nucleic acid with one or more compounds suspected of possessing
 CC binding affinity for the polypeptide or nucleic acid molecule, and
 CC selecting a compound that binds to AD or its nucleic acid.
 CC The polypeptides, nucleic acids, vector, ligand, or compound are
 CC useful in therapy or diagnosis of disease such as a cardiovascular
 CC disease (including atherosclerosis, ischaemia, restenosis, reperfusion
 CC injury, sepsis), a haematological disease (such as leukaemia), a blood
 CC clotting disorder (such as thrombosis), cancer (including lung,
 CC prostate, breast, colorectal and brain tumours, metastasis), an
 CC inflammatory disease such as rhinitis, a gastrointestinal disease
 CC (including inflammatory bowel disease, ulcerative colitis, Crohn's
 CC disease), a respiratory disease (including asthma, chronic obstructive
 CC pulmonary disease (COPD), respiratory distress syndrome, pulmonary
 CC fibrosis), immune disorders (including autoimmune diseases, rheumatoid
 CC arthritis, transplant rejection), allergy, liver diseases such as
 CC cirrhosis, endocrine diseases such as diabetes, bone diseases such as
 CC osteoporosis, neurological diseases (including stroke, multiple
 CC sclerosis, spinal cord injury), burns and wound healing, bacterial
 CC infection, particularly Mycobacterium tuberculosis infection, or virus
 CC infection. The present sequence is the adhesion molecule protein, AD3.

XX Sequence 1441 AA;

Query Match 8.0%; Score 85.5; DB 23; Length 1441;

Best Local Similarity 22.0%; Pred. No. 11;
 Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

OY 20 IHYDKI--TKEINKATIDDAHMAIEOSETIDPMK-----VPDHADKPERVH-----GIYDF 67

DB 343 LHRDTTVOTLTLOPSVKDGLIYEDSPYKAVKGLIIVLVADEADKAPVTNCTLKTLVE- 401

OY 68 KGEIARL--RNIEARGLKOMKROGDANVKEGEGYKAH-----LLIGVHDDIVSMEDYLA 119

DB 402 NCGMILADRRITVA-----NSANVNGREVVYIHRDFRNIYLANRGPPLGDNDF 452

OY 120 YKLGDLHPTTHYISD-----IDGFVALS--LEISDEGNITMWS 156

DB 453 GTLGDIIF-SCHAVDNKPRHSELEMLRQYGNVPEPIQLKLVAAFGELRSIADGIIIN-P 510

OY 157 FEVROFANVNVNHIGSLIDPIFGVLSVDLTAIFQ-DYVRKEMTKYL 202

DB 511 YSTREVVNIYVHLQKF---PTEG-LSSVVRNVEFDSYNDMDRETL 552

RESULT 27

AAE10924

ID AAE10924 standard. Protein: 2780 AA.

XX AAE10924;

DT 18-DEC-2001 (first entry)

XX Mouse monogenic audiogenic seizure-susceptible-1 (mass1) protein.

XX Mouse; monogenic audiogenic seizure-susceptible-1 gene; mass1;

XX transgenic animal; genetic abnormality; seizure.

OS Mus musculus.

XX Key

FT Domain

Location/Qualifiers

24..58

XX	Sequence		660 AA;		
XX	Query Match	8.1%; Score 86.5; DB 22:	Length 660;		
CC	Best Local Similarity	22.2%;	Pred. No. 2.8;		
CC	Matches	36; Conservative	31; Mismatches	50; Indels	45; Gaps
OY	1	MKKLLLAANAAVAVASADPIHYDKITEINKAIDDAIAIEQSEETDPMKV-----	51		
Dd	354	I FSLIMIALVSFVAAMFGNKEYEETDPVIKSYKEA----EQIFNNKNILGKISRSYS	409		
OY	52	-----PDHADKPER--HNGIYDFPG-----ELAMRNIRARGIKOM	84		
Dd	410	KYPENELIKTTPENGGERVEDSDVDVISKGPEKVPKPNIPLGEALOKLSLGLDV	469		
OY	85	KROGDANYKGEIGI-----VKAHLLICVHDIVSMEXDLAK	121		
Dd	470	KIEKYNNNQAPKGYIANOSVTANEINAIHDSNIKLYESLOIK	511		
RESULT 24					
AU36586	ID	AAU36586 standard; Protein; 664 AA.			
AC	AAU36586;				
DX	14-FEB-2002	(first entry)			
DE	Staphylococcus aureus cellular proliferation protein #756.				
KW	Antisense: prokaryotic cellular proliferation protein;				
KS	antibiotic; antibacterial; drug design.				
OS	Staphylococcus aureus.				
PN	MO200170955-A2.				
PP	27-SEP-2001.				
PR	21-MAR-2001; 2001WO-US09180.				
XX	21-MAR-2000; 2000US-191078P.				
XX	23-MAY-2000; 2000US-206848P.				
XX	26-MAY-2000; 2000US-207727P.				
XX	23-OCT-2000; 2000US-242578P.				
XX	27-NOV-2000; 2000US-253625P.				
XX	22-DEC-2000; 2000US-257931P.				
XX	16-FEB-2001; 2001US-269308P.				
PA	(ELIT-) ELITRA PHARM INC.				
PI	Hasselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;				
PT	Yamamoto RT, Xu HH;				
DR	N-PSDB; AAS54445.				
XX	WPI; 2001-611495/70.				
XX	New polynucleotides for the identification and development of				
XX	antibiotics, comprise sequences of antisense nucleic acids -				
XX	Example 3; Seq ID No 12179; 511bp; English.				
CC	The invention relates to antisense inhibitors of genes essential to				
CC	prokaryotic cellular proliferation, their use in identifying the				
CC	genes themselves and the encoded proteins. The prokaryotes used are				
CC	<i>Escherichia coli</i> , <i>Staphylococcus aureus</i> , <i>Salmonella typhi</i> , <i>Klebsiella</i>				
CC	<i>pneumoniae</i> , <i>Pseudomonas aeruginosa</i> and <i>Enterococcus faecalis</i> . The				
CC	invention is also useful for the identification of potential new targets				
CC	for antibiotic development. The antisense nucleic acids can also be used				
CC	to identify proteins used in proliferation, to express these proteins,				
CC	and to obtain antibodies capable of binding to the expressed proteins.				
CC	The proteins can be used to screen compounds in rational drug discovery				

	CC	programmes. The antisense nucleic acid sequence is also useful to screen
	CC	for homologous nucleic acids which are required for cell proliferation in
	CC	a wide variety of organisms. The present sequence represents an
	CC	essential prokaryotic cellular proliferation protein.
	CC	Note: The sequence data for this patent did not form part
	CC	of the printed specification, but was obtained in electronic
	CC	format directly from WIPO at
	CC	ftp.wipo.int/pub/published_pct_sequences.
	XX	
SQ	Sequence	664 AA;
	Query Match	8.1%; Score 86.5; DB 22; Length 664;
	Best Local Similarity	22.2%; Pred. No. 2.8;
MATCHES	Matches	36; Conservative 31; Mismatches 50; Indels 45; Gaps 5.
OY	1 MMKFLIIAAVFAVASAPRHHDKITEENKAIDAIAIFQSSEITDPMKV-----51	: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
Dd	355 IFSLLMIALVSRYAAMMGCKNYEELPDVIYGSKVKLA---EDIPFKNNIKLKGISRSYSDD410	:::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
OY	52 -----PDAHKFER--HGVIVDFKG-----ELARNINIEARGLKM 84	:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
Dd	411 KYPEKEIITKTPTNGTGERBERGDSVDVVISGCEKYKMPNVIGLPREQNALQTKLSGLKV 470	: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
OY	85 KRQGSDANVGEGEI----YKAHLILGYHDIDIVSNEYDLAK 121	: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :
Dd	471 KIEKYYNQAEPKGYIANOSVTANTEIATHSNIKLYESLGIR 512	:: :: :: :: :: :: :: :: :: :: :: :: :: :: :
RESULT 25		
ID	AAG82219 standard; Protein; 312 AA.	
AC	AAG82219;	
DT	03-SEP-2001 (first entry)	
DE	S. epidermidis open reading frame protein sequence SEQ ID NO:1532.	
KW	Staphylococcus epidermidis SRI strain; infection; diagnosis;	
XX	vaccination; endocarditis.	
OS	Staphylococcus epidermidis.	
PN	WO200134809-A2.	
Pd	17-MAY-2001.	
PF	09-NOV-2000; 2000WO-US30782.	
PR	09-NOV-1999; 99US-O164258.	
PA	(GLAX) GLAXO GROUP LTD.	
L1	kimmerly WJ;	
DR	WP1: 2001-316495/33.	
N-PSDB	; AAH53069.	
FT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,	
Cc	useful for vaccinating against infections, e.g. endocarditis -	
claim	18; Page 433; 2188pp; English.	
CC	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides	
CC	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.	
CC	(I) and (II) can have antibacterial activity and therefore can be used	
CC	in vaccination. The nucleic acids (I) may be used to produce the	
CC	S. epidermidis polypeptides (II) via the production of vectors	
CC	containing them which are used to produce host cells which express the	
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be	
CC	used to vaccine subjects and to raise antibodies against the bacteria.	
CC	The polypeptides may also be used to assay for other inhibitors of their	
CC	activity and therefore identify compounds that may be used for the	

Db 21 DKFSFDLKGENVKAMDIAIATKVGECVCHITCKPEYAVGSAGSPKIPPNATLVEE--V 78
 QY 63 GIYDFKGEGLAMRNIEARGKQKROGDANVKGEGIVKAKHLIGVHDDIYSMEYDLAYKL 122
 Db 79 ELFEFEGEDLTEREEDGCIIRIOTRGEGYAKPNEGAIVEYALGYYKDKLFDQRELRPEI 138
 QY 123 GD 124
 Db 139 GE 140

RESULT 18 AAM54038

ID AAM54038 standard; Protein; 459 AA.

AC AAM54038;

DT 06-AUG-1998 (first entry)

DE Human FK506 binding protein, FKBP52.

KM FK506 binding protein; FKBP52; human; immune response regulator;

KW immunosuppressant; steroid hormone receptor transformation.

OS Homo sapiens.

PN US5763590-A.

PD 09-JUN-1998.

PF 09-NOV-1994; 94US-0336618.

PR 16-OCT-1992; 92US-0963325.

PR 11-OCT-1991; 91US-0777752.

PR 29-MAR-1994; 94US-0218989.

PR 09-NOV-1994; 94US-0336618.

PA (VERT-) VERTEX PHARM INC.

PI Harding MW, Livingston DJ, Peattie DA;

DR WPI: 1998-347419/30.

DR N-PSDB; AAV24010.

PT DNA sequence encoding human FK506-binding protein - and recombinant

PT DNA molecule containing it

PS Example 5; Fig 3; 28pp; English.

XX This sequence is 52 kD human FK506 binding protein, referred to as

CC FKBP52, of the invention. The FKBP52 protein plays a key role in

CC regulating immune responses. FKBP52 may be useful for mediating steroid

CC hormone receptor transformation. The DNA may be used to screen for new

CC immunosuppressants, and in assays for metabolites in samples from

CC individuals taking immunosuppressants. The DNA may also be used in

CC assays for identifying natural intracellular rapamycin-like or FK506

CC like substances, and in assays for identifying natural intracellular

CC substrates that are potential targets for other immunosuppressants.

XX Sequence 459 AA:

Query Match 8.3%; Score 88.5; DB 19; Length 459;

Best Local Similarity 24.6%; Pred. No. 1;

Matches 30; Conservative 25; Mismatches 50; Indels 17; Gaps 3;

QY 18 DPHYDKTEBEINKAIDDAIAIEOSEI-----DPKKVDPHAD-KERRHV 62

Db 75 DKFSFDLKGENVKAMDIAIATKVGECVCHITCKPEYAVGSAGSPKIPPNATLVEE--V 132

QY 63 GIYDFKGEGLAMRNIEARGKQKROGDANVKGEGIVKAKHLIGVHDDIYSMEYDLAYKL 122

Db 133 ELFEFEGEDLTEREEDGCIIRIOTRGEGYAKPNEGAIVEYALGYYKDKLFDQRELRPEI 192

QY 123 GD 124
 Db 193 GE 194

RESULT 19 AAB04108

ID AAB04108 standard; Protein; 436 AA.

AC AAB04108;

DT 11-APR-2001 (first entry)

DE yphc protein of Streptococcus pneumoniae (GTP binding protein).

KM yphc; GTP binding protein; antibody; treatment; infection;

KW bacteraemia; otitis media; conjunctivitis; pneumoniae;

OS Streptococcus pneumoniae.

PN WO200068427-A1.

PD 16-NOV-2000.

PF 02-MAY-2000; 2000WO-US11894.

PR 07-MAY-1999; 99US-0307003.

PA (SMIR) SMITHKLINE BEECHAM CORP.

PA (SMIR) SMITHKLINE BEECHAM PLC.

PI Zalcain M, Biswas S, Burnham MKR, Warren PV, Ingraham KA;

PI Chalke AF, So CY, Holmes DJ, Warren RL, Van Horn S, Trainl CM;

DR WPI: 2000-687653/67.

DR N-PSDB; AAA54516.

PT Streptococcus pneumoniae yphc protein and DNA sequence, useful for

PT treating infections, meningitis, and bacteraemia

PS Claim 1; Page 37-38; 39pp; English.

XX A new nucleotide is described which encodes a 436 residue

CC Streptococcus pneumoniae yphc (GTP binding protein). The DNA

CC sequence can be used to transform a host cell to produce

CC the protein and the products can be used to treat bacterial

CC infections (especially Streptococcus pneumoniae infections, and

CC Helicobacter pylori infections), otitis media, conjunctivitis,

CC pneumoniae, bacteraemia, meningitis, sinusitis, pleural empyema, and

CC endocarditis. The encoded protein and its peptide fragments can be

CC used to generate antibodies.

XX Sequence 436 AA:

Query Match 8.2%; Score 88; DB 21; Length 436;

Best Local Similarity 21.8%; Pred. No. 1.1;

Matches 46; Conservative 34; Mismatches 89; Indels 42; Gaps 8;

QY 30 NKAIDDAIAIEOSEITIDPKVDPHADKFERHGVDFKGEGLAMRNIEARGKQKROGD 89

Db 21 NRIAGRISIVEDVEGVTDRIYATGEMLRNRSFSMIDTGG---IDVDAPFMEQIKHQA 77

QY 90 AN-----YKGEIGV-----AHLLIGVHDDIY-----SMEYDL---AYK 121

Db 78 IMEADVIVFVYSGKGTDADEYVARKLYTHKPVILAVNKVDNPEKNDIYDFYALG 137

QY 122 LGLDHTTIV---ISDIDFVVALSLEISDGNITMTSFEVROFANV-----VNHIGGL 172

Db 138 LGEPLPISSVHGICGVDVAIDVENLPNEYEENPVYIKFSLIGRNVGKSSLIINALGE 197

QY 173 S--ILDPIFGVLSDVLTALFIDTVRKEMTKV 201

DE Novel human diagnostic protein #25603.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSE INC.
XX
PI Drmanac RT, Liu C, Tang YT;
PI
DR WPI: 2001-639362/73.
DR N-PSDB: AAS89799.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 55971; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1164 AA;
XX
Query Match 8.3%; Score 89; DB 22; Length 1164;
Best Local Similarity 21.8%; Pred. No. 3.4;
Matches 51; Conservative 41; Mismatches 80; Indels 62; Gaps 10;
QY 6 LIAAFAVAVNSAD-----PIH--YDKTTEIN-----KAIDDAIAAIEQS 43
DB 217 VLAIPAFILINSVIGLILPAPFPHIRISALLEMKNTKFGSVKALDNCRLNAG 276
QY 44 ETID-----PMKVPDADKFEFRHVGIVDFKGLAMRIEARGLKOMKROGDANY 92
DB 277 EIVSLGEMSGSKTLMKVLGCIYPHGSEGEIIFAGE-----EIQASHIDTERRKGIATI 332
QY 93 KGEEGIVK-----AHLILIG--VHDDIVSMEXYDLAYKLGDLAPHTTVISDIQDFVALSL 144
DB 333 HQGLALVKELTVELNIFLGNEITHNGI--MDYDL-----MTLRCKQLLAQVSL 378
QY 145 EISDE---GNITMTSEVROFANVNHIGSLIDPIFGVLSVYLAIFODIYR 195
DB 379 SISPTRVGDLGGOOLVEIAKALNKQVRLTILDEPTASTTQETSILDIIR 432

RESULT 17
AAB58737
ID AAB58737 standard; Protein; 405 AA.
XX
AC AAB58737;
XX
DT 27-MAR-2001 (first entry)
XX
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 445.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neutrotropic; neuroprotective; antiviral; antiasthmatic; hepatotropic;
KW antidiabetic; antiinflammatory; antilucer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW Addison's disease; allergy; autoimmune hemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMANA GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
PI
DR WPI: 2000-611515/58.
DR N-PSDB: AAF21640.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer. Immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 11; Page 872-874; 1299pp; English.
XX
CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neutrotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antilucer; vulnerary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiac activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune hemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 405 AA;
XX
Query Match 8.3%; Score 88.5; DB 21; Length 405;
Best Local Similarity 24.6%; Pred. No. 0.84;
Matches 30; Conservative 25; Mismatches 50; Indels 17; Gaps 3;
QY 18 DPHYDKTTEINKAIDDAIAAIEQSETI-----DPKKVPDAD-KFERHY 62

XX This sequence corresponds to a Plasmodium falciparum strain K1
 CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
 CC gene sequence was isolated by screening a P. falciparum strain 79/96
 CC library with serum from a missionary treated by prophylaxis (for strain
 CC T6/96 see FR9101286). Of 20 clones isolated, clone 7295 was used to
 CC screen a library generated from Thai strain K1. One clone contained a
 CC 6.85 kb insert including the genomic sequence AAT78867. The gene
 CC comprises a 1.8 kb region encoding 3 major blocks of tetrapeptide
 CC repeats (especially the amino acid sequence VEEV, VEEN, VEEL, VAPS, VAPT,
 CC etc) and a 3' hydrophobic region corresponding to a
 CC glycosyl-phosphatidyl- inositol membrane anchoring sequence. The
 CC invention relates to new polypeptides of at least 10 amino acids derived
 CC from the LSA-3 protein with the exception of the peptides AAM24791-4.
 CC The LSA-3 peptides can be used to raise antibodies and as vaccines for
 CC immunotherapy of malaria.

SQ Sequence 1786 AA;

Query Match 8.4%; Score 90; DB 18; Length 1786;
 Best Local Similarity 23.3%; Pred. No. 4.8;
 Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;

OY 22 YDKITEINKAIDDAIAIEQSETIDPMKVPDHADKFERHVGIVDFKGLAMRNIEARGL 81
 DB 979 FNTVLDRKVEETVIEISGESELENNE---MDKAFSEIFDNVKGIQENLITGMFRSIETSIY 1034
 OY 82 KOMKROGDANKVGECEGIVAHLLIGVHDDIVSMEDYLAVKLGDLAFTTTHVISDIDQFVYA 141
 DB 1035 IOSEKRVVDLN---ENVVSSIL---DNIENKKEGLLNKLNEISSTEGVOETVTEHV-- 1083
 OY 142 LSLKISDEGNITMTSFEV---ROFANVVNHIGLS---IIDPIFGVLSDVLTAA--IFQ 191
 DB 1084 -----EONV-YVDVDVPAKMDQFIGIILNEAGLKEKMFNLEDVKSSEDVITVEIND 1135
 OY 192 DTVRKEMTVLAPAFKRELEKN 213
 DB 1136 EPVQKEVEKETVSIIE-EEMEEN 1156

RESULT 15
 AA096699
 ID AAU96699 standard; Protein; 1787 AA.
 AC AAU96699;
 XX
 DT 30-JUL-2002 (first entry)
 XX
 DE Plasmodium falciparum liver stage antigen-3 (LSA-3) #1.
 XX
 KW Liver stage antigen-3; LSA-3; vaccine; Th1-inducing adjuvant;
 KW malaria parasite; malaria; protein-specific cytotoxic T cell response;
 KW gamma-interferon.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FT Region 1..278
 FT /label= NR-A
 FT /note= Non-repeated_region_A
 FT Region 46..63
 FT /label= Hydrophobic_region_1
 FT 177..201
 FT /note= "Peptide used for chimpanzee immunisation"
 FT 198..223
 FT /note= "Peptide used for chimpanzee immunisation. Can
 be used as a palmitoyl-conjugated lipopeptide"
 FT 223..278
 FT /label= Conserved_repeat_region
 FT 279..819
 FT /label= Polymorphic_repeat_region
 FT 820..1535
 FT /label= Non-repeated_region_B

FT Region 1536..1577
 FT /label= Conserved_repeat_region
 FT Region 1578..1786
 FT /label= Non-repeated_region_C
 FT Region 1662..1694
 FT /label= Hydrophobic_region_2

PN EP1201250-A1.
 XX
 PD 02-MAY-2002.
 XX
 XX 25-OCT-2000; 2000EP-0203724.
 PF 25-OCT-2000; 2000EP-0203724.
 XX
 PR 25-OCT-2000; 2000EP-0203724.
 XX
 PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA (INSP) INST PASTEUR.
 XX
 PI Cohen J, Druilhe P;
 XX
 DR WPI: 2002-373883/41.
 DR N-PSDB; ABK50883.
 XX
 XX Vaccine for treatment or prevention of malaria, comprises liver stage
 PT antigen and adjuvant that induces Th1 response
 PS Disclosure: Page 17; 56pp; English.

CC The invention describes a vaccine comprising a Th1-inducing adjuvant (I)
 CC and a protective liver-stage antigen (LSA), or its immunogenic fragment,
 CC from a human malaria parasite, where LSA is a fragment of LSA-3, then (I)
 CC is not montanide. The vaccine is useful for treating or preventing
 CC malaria, specifically where caused by Plasmodium falciparum. The vaccine
 CC is formulated in SBA52 adjuvant, an oil-in-water emulsion containing QS21
 CC and 3D-MPL (de-O-acetylated monophosphoryl-lipid A, the combination of
 CC which results in strong induction of a circumsporozoite protein-specific
 CC cytotoxic T cell response, which is not generally induced by vaccines
 CC based on recombinant proteins, and synergistically increases production
 CC of gamma-interferon. This is the amino acid sequence of the liver-stage
 CC antigen-3 (LSA-3) from the Plasmodium falciparum strain K1 used to
 CC develop the recombinant proteins for the vaccine.

SQ Sequence 1787 AA;

Query Match 8.4%; Score 90; DB 23; Length 1787;
 Best Local Similarity 23.3%; Pred. No. 4.8;
 Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;

OY 22 YDKITEINKAIDDAIAIEQSETIDPMKVPDHADKFERHVGIVDFKGLAMRNIEARGL 81
 DB 979 FNTVLDRKVEETVIEISGESELENNE---MDKAFSEIFDNVKGIQENLITGMFRSIETSIY 1034
 OY 82 KOMKROGDANKVGECEGIVAHLLIGVHDDIVSMEDYLAVKLGDLAFTTTHVISDIDQFVYA 141
 DB 1035 IOSEKRVVDLN---ENVVSSIL---DNIENKKEGLLNKLNEISSTEGVOETVTEHV-- 1083
 OY 142 LSLKISDEGNITMTSFEV---ROFANVVNHIGLS---IIDPIFGVLSDVLTAA--IFQ 191
 DB 1084 -----EONV-YVDVDVPAKMDQFIGIILNEAGLKEKMFNLEDVKSSEDVITVEIND 1135
 OY 192 DTVRKEMTVLAPAFKRELEKN 213
 DB 1136 EPVQKEVEKETVSIIE-EEMEEN 1156

RESULT 16
 ABG25612
 ID ABG25612 standard; Protein; 1164 AA.
 XX
 AC ABG25612;
 XX
 DT 18-FEB-2002 (first entry)
 XX

PS Claim 1; Page 61; 155pp; English.

XX The present invention describes polypeptides (I), designated POLXX

CC polypeptides (e.g. POLY1, POLY2, etc.) and the polynucleotide

CC sequences (II) encoding them. POLY1-4 are members of the gamma

CC aminobutyric acid (GABA) receptor family; POLY5-8 are members of the

CC epidermal growth factor (EGF) family; POLY9-11 are members of the

CC complement receptor family; POLY12 is a member of the haematopoietic

CC stem and progenitor cell (HSPC) family; POLY13 is a member of the

CC sulphotransferase family; POLY14-16 are members of the syntaxin family;

CC and POLY17 is a member of the prohibitin family. (I) and (II) can have

CC antidepressant, cerebroprotective, antiparkinsonian, nootropic, relaxant,

CC anticonvulsant, cardioprotective, neuroprotective, antialcoholic, cardiant,

CC tranquiliser and antiarrhythmic activities. (I) and (II) can be used

CC for treating or preventing a POLXX-associated disorder in humans as a

CC therapeutic in the manufacture of a medicament for treating a syndrome

CC associated with a human disease selected from a POLXX-associated

CC disorder, for treating a pathological state in a mammal, especially

CC patients suffering from, e.g., psychiatric and medical conditions,

CC depression, stroke, Parkinson's disease, Huntington's disease,

CC Tourette's syndrome, amyotrophic lateral sclerosis, head trauma,

CC Alzheimer's disease, alcoholism, vigilance, anxiety, muscle tension,

CC epileptogenic activity and memory functions, cardiomyopathy and

CC arrhythmogenic right ventricular dysplasia. The present sequence

CC represents POLY16.

XX

SQ Sequence 294 AA:

Query Match 9.0%; Score 96; DB 23; Length 294;

Best Local Similarity 23.8%; Pred. No. 0.084;

Matches 44; Conservative 31; Mismatches 58; Indels 52; Gaps 10;

OY 27 EEINKAIDDAIAIEOSEETID-PMKV-----PDHADKFEHNGIVDF-----KGE-70

DB 103 EYINRSLNDLVKRVKSEVENGPSSVTRIKSQHAMF-RHQOQIMFYINDTIANKQEK 161

OY 71 ---LAMRNIEARG-----LKQMKRQGDANVKGEE---GIVKAHL-LIGVHDDIVS 113

DB 162 CKTFILRLOLVAGKEMSEEDVNDMLHQKWEKVENESLTFEINITKQSLSEIQRHKELYN 221

OY 114 MEYDLAKYKGLDHPHTHVISDIODFYVALSLSEISDEG---NITMPSFEVROKQANVNH 169

DB 222 LE-----NQIKDLRDLFTQISLVEEGEGESINTEIVNSTKEYVNMNTEK 267

OY 170 GGLSI 174

DB 268 FGLAV 272

RESULT 9

AAU19614

ID AAU19614 standard; Protein: 314 AA.

XX

AC AAU19614:

XX

DT 04-DEC-2001 (first entry)

XX

DE Human diagnostic and therapeutic polypeptide (DITHP) #200.

XX

KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;

KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;

KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;

KW respiratory disorder.

XX

OS Homo sapiens.

XX

PN WO200162927-A2.

XX

PD 30-AUG-2001.

XX

PF 21-FEB-2001; 2001WO-US06059.

XX

PR 24-FEB-2000; 2000US-0184693.

PR 24-FEB-2000; 2000US-0184697.

PR 24-FEB-2000; 2000US-0184698.

PR 24-FEB-2000; 2000US-0184768.

PR 24-FEB-2000; 2000US-0184769.

PR 24-FEB-2000; 2000US-0184770.

PR 24-FEB-2000; 2000US-0184771.

PR 24-FEB-2000; 2000US-0184772.

PR 24-FEB-2000; 2000US-0184773.

PR 24-FEB-2000; 2000US-0184774.

PR 24-FEB-2000; 2000US-0184776.

PR 24-FEB-2000; 2000US-0184777.

PR 24-FEB-2000; 2000US-0184797.

PR 24-FEB-2000; 2000US-0184813.

PR 24-FEB-2000; 2000US-0184837.

PR 24-FEB-2000; 2000US-0184841.

PR 24-FEB-2000; 2000US-0185213.

PR 24-FEB-2000; 2000US-0185216.

PR 12-MAY-2000; 2000US-0203785.

PR 15-MAY-2000; 2000US-0204226.

PR 16-MAY-2000; 2000US-0204525.

PR 16-MAY-2000; 2000US-0204821.

PR 16-MAY-2000; 2000US-0204908.

PR 16-MAY-2000; 2000US-0205232.

PR 17-MAY-2000; 2000US-0204815.

PR 17-MAY-2000; 2000US-0204863.

PR 17-MAY-2000; 2000US-0205221.

PR 17-MAY-2000; 2000US-0205285.

PR 17-MAY-2000; 2000US-0205286.

PR 17-MAY-2000; 2000US-0205287.

PR 17-MAY-2000; 2000US-0205323.

PR 17-MAY-2000; 2000US-0205324.

XX

PA (INCYTE) INCYTE GENOMICS INC.

XX

PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;

PI Chen A, D'Sa SA, Amshy S, Dahl CR, Dam JC, Daniels SE;

PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;

PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;

PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;

PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;

XX

DR WPI: 2001-502867/55.

DR N-PSDB: AAS31185.

XX

PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.

PT enzymes, hormones and receptors, useful in diagnostics and therapeutics

PT

XX

PS Claim 27; Page 516; 522pp; English.

XX

PS The invention relates to polynucleotides (I) encoding diagnostic and

CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,

CC and proteins involved in growth and development and receptors. (I) and

CC (II) may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate DITHP expression. For example, (I) and

CC (II) may be used to treat disorders associated with decreased polypeptide

CC expression by rectifying mutations or deletions in a patient's genome,

CC that affect the activity of the DITHPs, by expressing inactive proteins

CC or supplementing the patient's own production of them. (I) and (II)

CC may be used to treat diseases, for example, cell proliferative disorder,

CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,

CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,

CC (I) may be used to produce the DITHPs, by inserting the nucleic acids

CC into a host cell and culturing the cell to express the protein. (I) and

CC its complementary sequences may also be used as DNA probes in diagnostic

CC assays to detect and quantitate the presence of similar nucleic acids in

CC samples, and therefore which patients may be in need of restorative

CC therapy. (II) may also be used as antigens in the production of

CC antibodies against DITHPs and in assays to identify modulators of DITHP

CC expression and activity. The anti-DITHP antibodies and antagonists may

CC also be used to down regulate expression and activity. The anti-DITHP

CC antibodies may also be used as diagnostic agents for detecting the

CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant

KW Human: POLYX; gamma aminobutyric acid receptor; GABA receptor;
 KW Epidermal growth factor; EGF; complement receptor; HSPC; syntaxin;
 KW haematopoietic stem and progenitor cell; sulphotransferase; prohibitin;
 KW antidepressant; cerebroprotective; antiparkinsonian; nootropic; relaxant;
 KW anticonvulsant; neuroleptic; neuroprotective; antialcoholic; cardiant;
 KW tranquilliser; antiarrhythmic; psychiatric; medical; depression; stroke;
 KW Parkinson's disease; Huntington's disease; Tourette's syndrome; anxiety;
 KW amyotrophic lateral sclerosis; head trauma; Alzheimer's disease;
 KW alcoholism; vigilance; muscle tension; epileptogenic; memory function;
 KW cardiomyopathy; arrhythmogenic right ventricular dysplasia.
 XX
 OS Homo sapiens.
 PN WO200179294-A2.
 PD 25-OCT-2001.
 PF 19-APR-2001; 2001WO-US12854.
 XX
 PR 19-APR-2000; 2000US-198293P.
 PR 20-APR-2000; 2000US-198645P.
 PR 25-APR-2000; 2000US-199476P.
 PR 26-APR-2000; 2000US-199860P.
 PR 26-APR-2000; 2000US-200023P.
 PR 26-APR-2000; 2000US-200025P.
 PR 09-JUN-2000; 2000US-210809P.
 PR 17-JUL-2000; 2000US-218591P.
 PR 11-AUG-2000; 2000US-224619P.
 PR 09-FEB-2001; 2001US-267673P.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Tauplier RJ, Vernet CAM, Fernandes E, Shinkets RA, Majumder K;
 PI Padigaru M, Colman SD, zerhusen BD, Szytek KA, Burgess CE, Liu X;
 XX
 DR WPI: 2002-017601/02.
 DR N-PSDB; ABA03884.
 PT
 PT New isolated polypeptides for treating a broad range of pathological
 PT states, e.g., depression, stroke, Parkinson's disease, Huntington's
 PT disease, Tourette's syndrome, amyotrophic lateral sclerosis, head
 PT trauma, and Alzheimer's -
 XX
 PS Claim 1; Page 59; 155pp; English.
 XX
 CC The present invention describes polypeptides (I), designated POLYX
 CC polypeptides (e.g. POLY1, POLY2, etc.) and the polynucleotide
 CC sequences (II) encoding them. POLY1-4 are members of the gamma
 CC aminobutyric acid (GABA) receptor family; POLY5-8 are members of the
 CC complement growth factor (EGF) family; POLY9-11 are members of the
 CC complement receptor family; POLY12 is a member of the haematopoietic
 CC stem and progenitor cell (HSPC) family; POLY13 is a member of the
 CC sulphotransferase family; POLY14-16 are members of the syntaxin family;
 CC and POLY17 is a member of the prohibitin family. (I) and (II) can have
 CC antidepressant, cerebroprotective, antiparkinsonian, nootropic, relaxant,
 CC anticonvulsant, neuroleptic, neuroprotective, antialcoholic, cardiant,
 CC tranquilliser and antiarrhythmic activities. (I) and (II) can be used
 CC for treating or preventing a POLYX-associated disorder in humans as a
 CC therapeutic in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease selected from a POLYX-associated
 CC disorder, for treating a pathological state in a mammal, especially
 CC patients suffering from, e.g., psychiatric and medical conditions,
 CC depression, stroke, Parkinson's disease, Huntington's disease,
 CC Tourette's syndrome, amyotrophic lateral sclerosis, head trauma,
 CC Alzheimer's disease, alcoholism, vigilance, anxiety, muscle tension,
 CC epileptogenic activity and memory functions, cardiomyopathy and
 CC arrhythmogenic right ventricular dysplasia. The present sequence
 CC represents POLY15.
 XX
 SO Sequence 294 AA;
 Query Match 9.0%; Score 96; DB 23; Length 294;

QY	27	SEINNAIDDAIAALAEQSEFTID-PMKV-----PDHAEFERHVGIVDF-----KGE-70
Db	103	EYIRNSLNDLYKEVKKSEVENGPSSVYRIKLSQNAAMF-RHFQOIMFYNTITIAKQEK 161
QY	71	---LAMNRIEARG-----LKQMKRGCDANYKGE-----GIYKHL-LIGYDDIYS 113
Db	162	CKTFPIRLQLEAVGAKMSEEDVDNMHQQCKWEVFNLSLTLEINIFTKAOLSEIQRRKEIVN 221
QY	114	MEYDANYKLGDLHPHTYHSIDPIODEVVALSLSEISDEG---NITMTSEVRQFANVNI 169
Db	222	LE-----NQIKDLRLDLFIQISLVEQGESINNEMTVMSTKEYVNNTKEK 267
QY	170	GGLSI 174
Db	268	FGIAV 272
RESULT 8		
AAAG68269		
ID	AAAG68269	standard; Protein; 294 AA.
AC	AAAG68269;	
XX		
DE	14-FEB-2002	(first entry)
XX		
DE	Human POLY16	protein sequence SEQ ID NO:32.
XX		
KM	Human; POLY16	gamma aminobutyric acid receptor; GABA receptor;
KM	epidermal growth factor; EGF; complement receptor; HSPC; syntaxin;	
KM	haematopoietic stem and progenitor cell; sulphotransferase; prothibitin;	
KM	antidepressant; cerebroprotective; antiparkinsonian; nootropic; relaxant;	
KM	anticonvulsant; neuroleptic; neuroprotective; antialcoholic; cardiact;	
KM	tranquilliser; antiarrhythmic; psychiatric; medical; depression; stroke;	
KM	Parkinson's disease; Huntington's disease; Tourette's syndrome; anxiety;	
KM	amyotrophic lateral sclerosis; head trauma; Alzheimer's disease;	
KM	alcoholism; vigilance; muscle tension; epileptogenic; memory function;	
KM	cardiomyopathy; arrhythmogenic right ventricular dysplasia.	
OS	Homo sapiens.	
XX		
PN	WO200179294-A2.	
XX		
PD	25-OCT-2001.	
XX		
PE	19-APR-2001;	2001WO-US12854.
XX		
PR	19-APR-2000;	2000US-198293P.
PR	20-APR-2000;	2000US-198645P.
PR	25-APR-2000;	2000US-199476P.
PR	26-APR-2000;	2000US-199880P.
PR	26-APR-2000;	2000US-200024P.
PR	26-APR-2000;	2000US-200025P.
PR	09-JUN-2000;	2000US-210809P.
PR	17-JUL-2000;	2000US-218591P.
PR	11-AUG-2000;	2000US-224610P.
PR	09-FEB-2001;	2001US-267673P.
PA		
XX	(CURA-)	CURAGEN CORP.
PI	Taupier RJ,	Vernet CAM, Fernandes E, Shinkets RA, Majumder K;
PI	Padigaru M,	Colman SD, Zerhusen BD, Spytek KA, Burgess CE, Liu X;
XX		
DR	WPI: 2002-017601/02.	
XX	N-PSDB; ABA03885.	
PT	New isolated polypeptides for treating a broad range of pathological	
PT	sites, e.g., depression, stroke, Parkinson's disease, Huntington's	
PT	disease, Tourette's syndrome, amyotrophic lateral sclerosis, head	
PT	trauma, and Alzheimer's	
XX		

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 260 AA:

Query Match 9.3%; Score 99; DB 22; Length 260;

Best Local Similarity 21.9%; Pred. No. 0.033; Mismatches 63; Indels 96; Gaps 13;

Matches 56; Conservative 41; Mismatches 63; Indels 96; Gaps 13;

QY 5 LIIAAVAVVAVSA-----DPHYDKTTERINKAIDDAIAEQS-----ETIDPM 49
DB 3 IFVALLAFVAVASASMGPIE---TQSSSTIVDVIGIKEQMPCGFTSVGLPPLAFL 58
QY 50 KVPDADKFERHVGIVDFKGEIAMRNIEARGLKQMKRGCDANVKGEGIVAKHLIGVHD 109
DB 59 RI-DHQD-----INI-----DSSVLKAKQGTIDHFRGLND 88
QY 110 -DI-----VSMEYDLAYKGLDLPHTTHYS-----DIODEFYV 140
DB 89 FIDIEKKNVAITSKYTKTFPRDQVNDYQDLSVLKKGFTINLIGAGHAKFAIKDWYI 148
QY 141 ALSEIS---DEGNITMFSFEVROFAMVYNHIGLSIIDPIGVLSDVLTALFODTRKE 197
DB 149 WQTMKYSLSGIVISGNLKLKSLFVR-----THLG--EVDEISGIIID-----GSINK 193
QY 198 MKKVLAPAFKRELEKN 213
DB 194 MNEYLAEVAELAINEN 209

XX Db

RESULT 6

AA68267

ID AAG68267 standard; Protein: 294 AA.

XX AC AAG68267;

XX DT 14-FEB-2002 (first entry)

XX DE Human POLY14 protein sequence SEQ ID NO:28.

XX KW Human: POLYX; gamma aminobutyric acid receptor; GABA receptor;
KW epidermal growth factor; EGF; complement receptor; HSPC; syntaxin;
KW haematopoietic stem and progenitor cell; sulphotransferase; prohibitin;
KW antidepressant; cerebroprotective; antiparkinsonian; nootropic; relaxant;
KW anticonvulsant; antiaerhythmic; neuropsychiatric; medical; depression; stroke;
KW Parkinson's disease; Huntington's disease; Tourette's syndrome; anxiety;
KW amyotrophic lateral sclerosis; head trauma; Alzheimer's disease;
KW alcoholism; vigilance; muscle tension; epileptogenic; memory function;
KW cardiomyopathy; arrhythmogenic right ventricular dysplasia;
KW chromosome 1.

XX OS Homo sapiens.

XX PN W0200179294-A2.

XX PD 25-OCT-2001.

XX PF 19-APR-2001; 2001WO-US12854.

XX PR 19-APR-2000; 2000US-198293P.

XX PR 20-APR-2000; 2000US-198645P.

XX PR 23-APR-2000; 2000US-199476P.

XX PR 26-APR-2000; 2000US-199880P.

XX PR 26-APR-2000; 2000US-200024P.

XX PR 09-JUN-2000; 2000US-210809P.

XX PR 17-JUL-2000; 2000US-218591P.

XX PR 11-AUG-2000; 2000US-224610P.

PR 09-FEB-2001; 2001US-267673P.

XX PA (CURA-) CURAGEN CORP.

XX PI Taupier RJ, Vernet CAM, Fernandes E, Shinkets RA, Majumder K;

XX PI Padigar M, Colman SD, Zernusen BD, Spytek KA, Burgess CE, Liu X;

XX WPI; 2002-017601/02.

XX N-PSDB; ABA03883.

XX PT New isolated polypeptides for treating a broad range of pathological

XX PT states, e.g., depression, stroke, Parkinson's disease, Huntington's

XX PT trauma, and Alzheimer's.

XX PS Claim 1; Page 57; 155pp; English.

XX CC The present invention describes polypeptides (I), designated POLYX
CC polypeptides (e.g. POLY1, POLY2, etc.), and the polynucleotide
CC sequences (II) encoding them. POLY1-4 are members of the gamma
CC aminobutyric acid (GABA) receptor family; POLY5-8 are members of the
CC epidermal growth factor (EGF) family; POLY9-11 are members of the
CC complement receptor family; POLY12 is a member of the haematopoietic
CC stem and progenitor cell (HSPC) family; POLY13 is a member of the
CC sulphotransferase family; POLY14-16 are members of the syntaxin family;
CC and POLY17 is a member of the prohibitin family. (I) and (II) can have
CC antidepressant, cerebroprotective, antiparkinsonian, nootropic, relaxant,
CC anticonvulsant, neuroleptic, neuropsychiatric, antialcoholic, cardiant,
CC tranquiliser and antiarrhythmic activities. (I) and (II) can be used
CC for treating or preventing a POLYX-associated disorder in humans as a
CC therapeutic in the manufacture of a medicament for treating a syndrome
CC associated with a human disease selected from a POLYX-associated
CC disorder, for treating a pathological state in a mammal, especially
CC patients suffering from, e.g., psychiatric and medical conditions,
CC depression, stroke, Parkinson's disease, Huntington's disease,
CC Tourette's syndrome, amyotrophic lateral sclerosis, head trauma,
CC Alzheimer's disease, alcoholism, vigilance, anxiety, muscle tension,
CC epileptogenic activity and memory functions, cardiomyopathy and
CC arrhythmogenic right ventricular dysplasia. The present sequence
CC represents POLY14, which was identified on chromosome 1.

XX XX Sequence 294 AA:

Query Match 9.0%; Score 96; DB 23; Length 294;

Best Local Similarity 23.8%; Pred. No. 0.084; Mismatches 58; Indels 52; Gaps 10;

Matches 44; Conservative 31; Mismatches 58; Indels 52; Gaps 10;

QY 27 EINKAIDDAIAIEQSEITD-PMKV-----PDHAKKERHVGIVDF-----KGE- 70

DB 103 EYINSLNDLVEVKSEVENPSSVWPRILKSOHAAAF-RHFOOIMFYNDTIAKOEK 161

QY 71 ---LWNRTEARG-----LKMKGCDANYKGE-----GIVKAHL--LLGVHDDIYS 113

DB 162 CFTFLIRQLEVAGKESSEVDNDMLHOGKWEVFNESLTFEITITRAQLSEIQRKELVN 221

QY 114 MEYDLAYKGLDLPHTTHYSIDQDFVALSLSEISDEG---NITMTSFEVROFAMVYNH 169

DB 222 LE-----NQIKDLRDLFIQISLVEEGESINNIEMTVNSTKEVNNTRKEK 267

QY 170 GGLST 174

DB 268 FGLAV 272

XX XX AAG68268

XX ID AAG68268 standard; Protein: 294 AA.

XX AC AAG68268;

XX DT 14-FEB-2002 (first entry)

XX DE Human POLY15 protein sequence SEQ ID NO:30.

CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitising patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC Tenbrio mollitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents the house dust mite (Dermatophagoides pteronyssinus) allergen
CC Der p 7.
CC
CC
XX
SQ Sequence 215 AA:
Query Match 90.1%; Score 962; DB 20; Length 215;
Best Local Similarity 85.9%; Pred. No. 4.7e-95;
Matches 183; Conservative 18; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKKELLIAAVAFVAVSADPIHYDKITEETENKAIDAIAIEQSEETIDPKVVDHADKFER 60
DB 1 MKKLLIIAAAFVAVSADPIHYDKITEETENKAIDAVAAIEKSETFDPKVPDHSDKFER 60
QY 61 HVGIVDFKGLAMRNIEARGLKQMKROGDANVKGEGIVKAHLIGVHDDIVSMEDLAY 120
DB 61 HIGIIDLKGLDMRNIOVRGLQMKRVGDANKSDEGVVKAHLIGVHDDIVSMEDLAY 120
QY 121 KLGDLHPTTHVYSIDIQDFVVALSLSEISDEGNITMTSFEVROFANVNHIGLSIIDPIFG 180
DB 121 KLGDLHPTTHVYSIDIQDFVVELSEVSEGNMTLTSFEVROFANVNHIGLSIIDPIFA 180
QY 181 VLSDLVLTAFODTVRKEMTKVLAPAFKRELEKN 213
DB 181 VLSDLVLTAFODTVRAEMTKVLAPAFKKELEERN 213
RESULT 4
AAR60575
ID AAR60575 standard; Protein: 215 AA.
XX
AC AAR60575;
XX
DT 01-APR-1995 (first entry)
XX
DE House dust mite allergen DerpVII cDNA.
XX
KW DerpVII allergen; antiallergic; allergy diagnosis.
XX
OS Dermatophagoides pteronyssinus.
XX
PN WO9420614-A.
XX
PD 15-SEP-1994.
XX
PE 11-MAR-1994; 94WO-AU00117.
XX
PR 12-MAR-1993; 93US-0031141.
XX
PR 22-JUN-1993; 93US-0081540.
XX
PA (CHIL-) INST CHILD HEALTH RES.
XX
PI Chua K, Thomas WR;
XX
DR WPI: 1994-303021/37.
XX
DR N-PSDB; AAO71400.
XX
PT New nucleic acid encoding specific dust mite allergens - and
PT related vectors, transformed cells, peptides and antibodies,
PT useful for desensitisation and diagnosis.

XX
PS Claim 7; Page 36-37; 67pp; English.
XX
XX DerpVII antigen is useful as antiallergic reagent for treating
CC sensitivity to house dust mite allergens. The DNA can be used
CC as a probe to detect the sensitivity of an individual to the
CC allergen.
XX
SQ Sequence 215 AA:
Query Match 89.8%; Score 959; DB 15; Length 215;
Best Local Similarity 85.4%; Pred. No. 9.9e-95;
Matches 182; Conservative 19; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKKELLIAAVAFVAVSADPIHYDKITEETENKAIDAIAIEQSEETIDPKVVDHADKFER 60
DB 1 MKKLLIIAAAFVAVSADPIHYDKITEETENKAIDAVAAIEKSETFDPKVPDHSDKFER 60
QY 61 HVGIVDFKGLAMRNIEARGLKQMKROGDANVKGEGIVKAHLIGVHDDIVSMEDLAY 120
DB 61 HIGIIDLKGLDMRNIOVRGLQMKRVGDANKSDEGVVKAHLIGVHDDIVSMEDLAY 120
QY 121 KLGDLHPTTHVYSIDIQDFVVALSLSEISDEGNITMTSFEVROFANVNHIGLSIIDPIFG 180
DB 121 KLGDLHPTTHVYSIDIQDFVVELSEVSEGNMTLTSFEVROFANVNHIGLSIIDPIFA 180
QY 181 VLSDLVLTAFODTVRKEMTKVLAPAFKRELEKN 213
DB 181 VLSDLVLTAFODTVRAEMTKVLAPAFKKELEERN 213
RESULT 5
ABB64121
ID ABB64121 standard; Protein: 260 AA.
XX
AC ABB64121;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 19155.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PERE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
DR N-PSDB; ABL08224.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 19155; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX Claim 5; Page 40-41; 67pp: English.
 PS DerivII antigen is useful as antiallergic reagent for treating
 CC sensitivity to house dust mite allergens.
 CC
 XX

SQ Sequence 213 AA:

Query Match 100.0%; Score 1068; DB 15; Length 213;
 Best Local Similarity 100.0%; Pred. No. 1.9e-106;
 Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKFLIAAFAVAVSADPIHYDKITEINKAIDDAIAIEOSETIDPKVPPHAKFER 60
 DB 1 MKKFLIAAFAVAVSADPIHYDKITEINKAIDDAIAIEOSETIDPKVPPHAKFER 60
 QY 61 HVGIVDFKGLAMRNIEARGLKQMKRGDANVGEIGVAAHLIGVHDDIVSMETDLAY 120
 DB 61 HVGIVDFKGLAMRNIEARGLKQMKRGDANVGEIGVAAHLIGVHDDIVSMETDLAY 120
 QY 121 KLGLDLPHTTHVISDIDFVVALSLEISDEGNITMTSFEVQFANVYVNHIGLSILDPPIFG 180
 DB 121 KLGLDLPHTTHVISDIDFVVALSLEISDEGNITMTSFEVQFANVYVNHIGLSILDPPIFG 180
 QY 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213
 DB 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213

RESULT 2
 AAY25592
 ID AAY25592 standard; protein; 213 AA.

XX AAY25592:

DT 30-SEP-1999 (first entry)
 XX
 DE D. farinae allergen Der f 7 protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;
 KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

OS Dermatophagoides farinae.

XX WO9934826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 52; 117pp: English.

CC This invention describes a novel method of desensitizing a patient to a
 CC polypeptide allergen and comprises administering to the patient a peptide
 CC derived from the allergen where restriction to a MHC Class II molecule
 CC possessed by the patient can be demonstrated for the peptide and the
 CC peptide is able to induce a late phase response in an individual who
 CC possesses the MHC Class II molecule. The methods can be used for

CC desensitizing patients to allergens present in e.g. grass, tree and weed
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
 CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
 CC Tenbrio mollitor beetle, mammals such as cat, dog, horse, cow, pig,
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
 CC produce immunological vaccines which may be used to prevent and/or treat
 CC conditions involving hypersensitivity to allergens. This sequence
 CC represents the Dermatophagoides farinae allergen Der f 7.

SQ Sequence 213 AA:

Query Match 100.0%; Score 1068; DB 20; Length 213;
 Best Local Similarity 100.0%; Pred. No. 1.9e-106;
 Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKFLIAAFAVAVSADPIHYDKITEINKAIDDAIAIEOSETIDPKVPPHAKFER 60
 DB 1 MKKFLIAAFAVAVSADPIHYDKITEINKAIDDAIAIEOSETIDPKVPPHAKFER 60
 QY 61 HVGIVDFKGLAMRNIEARGLKQMKRGDANVGEIGVAAHLIGVHDDIVSMETDLAY 120
 DB 61 HVGIVDFKGLAMRNIEARGLKQMKRGDANVGEIGVAAHLIGVHDDIVSMETDLAY 120
 QY 121 KLGLDLPHTTHVISDIDFVVALSLEISDEGNITMTSFEVQFANVYVNHIGLSILDPPIFG 180
 DB 121 KLGLDLPHTTHVISDIDFVVALSLEISDEGNITMTSFEVQFANVYVNHIGLSILDPPIFG 180
 QY 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213
 DB 181 VLSDVLTALFQDTPVRKEMTKVLAPAFKRELEKN 213

RESULT 3
 AAY25586
 ID AAY25586 standard; protein; 215 AA.

XX AAY25586:

DT 30-SEP-1999 (first entry)
 XX
 DE D. pteronyssinus allergen Der p 7 protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;
 KM allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
 KM chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
 KM screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
 KM cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
 KM mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

OS Dermatophagoides pteronyssinus.

XX WO9934826-A1.

XX 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

XX 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX Example 6; Page 51; 117pp: English.

CC This invention describes a novel method of desensitizing a patient to a

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 09:58:34 ; Search time 37 Seconds
(without alignments)
767.091 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068
Sequence: 1 MMKFLIAAVFAVAVSADPL.....VRKEWTKYLAPAFKRELEKN 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1068	100.0	213	15	House dust mite al
2	1068	100.0	213	15	D. farinose allerg
3	962	90.1	215	20	D. pteronyssinus a
4	959	89.8	215	15	House dust mite al
5	99	9.3	260	22	Drosophila melanog
6	96	9.0	294	23	Human POLY14 prote
7	96	9.0	294	23	Human POLY15 prote
8	96	9.0	294	23	Human POLY16 prote
9	96	9.0	314	22	Human diagnostic a
10	96	9.0	314	23	Human MDP1 Seq ID

11	93	8.7	436	23	ABP27826
12	91.5	8.6	270	22	AAU66088
13	90	8.4	1558	21	AAAB18324
14	90	8.4	1786	18	AAW24790
15	90	8.4	1787	23	AAU96699
16	89	8.3	1164	22	ABG25612
17	88.5	8.3	405	21	AAAB58737
18	88.5	8.3	459	19	AAW54038
19	88	8.2	436	21	AAAB04108
20	88	8.2	436	21	AAAB1730
21	87	8.1	436	23	ABP27827
22	87	8.1	1640	23	ABBS4727
23	86.5	8.1	660	22	AAU33974
24	86.5	8.1	664	22	AAU36586
25	86	8.1	312	22	AAAG82219
26	85.5	8.0	1441	23	AAU75885
27	85.5	8.0	2780	22	AAE10924
28	85	8.0	878	23	ABBS4705
29	84.5	7.9	293	14	AAAB35439
30	84.5	7.9	436	22	AAU58103
31	84.5	7.9	817	22	AAAG81478
32	84.5	7.9	817	22	AAAG82217
33	84.5	7.9	823	23	ABP35236
34	84.5	7.9	1396	20	AAV36871
35	84.5	7.9	1887	22	ABBS8245
36	84	7.9	306	23	ABP27046
37	84	7.9	410	20	AAV5996
38	84	7.9	775	23	ABBS3885
39	83.5	7.8	412	22	AAAG89772
40	83	7.8	845	23	AAU71861
41	82.5	7.8	1060	23	ABBS3805
42	82.5	7.7	548	18	AAAB1678
43	82.5	7.7	2486	22	AAU32848
44	82	7.7	277	22	AAAG82447
45	82	7.7	362	23	ABP39212

ALIGNMENTS

RESULT 1	
AAAR60576	
ID	AAAR60576 standard; Protein; 213 AA.
AC	AAAR60576;
DT	01-APR-1995 (first entry)
XX	
DE	House dust mite allergen DerfVII.
XX	
KW	DerfVII allergen; antiallergic; allergy diagnosis.
XX	
OS	Dermatophagoides farinae.
XX	
PN	WO9420614-A.
XX	
PD	15-SEP-1994.
XX	
PF	11-MAR-1994; 94WO-AU00117.
XX	
PR	12-MAR-1993; 93US-0031141.
PR	22-JUN-1993; 93US-0081540.
XX	
PA	(CHIL-) INST CHILD HEALTH RES.
XX	
PI	Chua K, Thomas WR;
XX	
DR	WPI: 1994-303021/37.
DR	N-PSDB; AAQ71401.
XX	
PT	New nucleic acid encoding specific dust mite allergens - and related vectors, transformed cells, peptides and antibodies, useful for desensitisation and diagnosis.
PT	

Streptococcus poly
Propionibacterium
Plasmodium falcipa
P. falciparum live
Plasmodium falcipa
Novel human diagno
Breast and ovarian
Human FK506 bindin
YpHc protein of St
Streptococcus pneu
Streptococcus poly
Lactococcus lactis
Staphylococcus aur
Staphylococcus aur
S. epidermidis ope
Human adhesion mol
Mouse monogenic au
Lactococcus lactis
B. burgdorferi str
Propionibacterium
S. epidermidis ope
S. epidermidis ope
Staphylococcus epi
Protein involved i
Drosophila melanog
Streptococcus poly
Human endometrium
Lactococcus lactis
C glutamicum prote
Leishmania antigen
Lactococcus lactis
Lawsonia intracell
Novel human secret
S. epidermidis ope
Staphylococcus epi


```

DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DE Hypothetical protein YPO1995.
GN YPO1995.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
NCBI_TaxID=632;
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltham D., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL: AJ414151; CAC90808.1; -.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 310 AA; 35163 MW; DBB4C4022BFP9B396 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 310;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 VLSDVLT 187
DB 278 VLSDVLT 284
|||||

RESULT 49
Q8Y307 PRELIMINARY; PRT; 313 AA.
AC Q8Y307;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Putative oxidoreductase protein (EC 1.1.-.-).
GN RSC0174 OR RS01045.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Trehault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646057; CAD13702.1; -.
DR InterPro: IPR003710; APDA.
DR Pfam: PF02558; APDA; 1.
KM Oxidoreductase; Complete proteome.
SQ SEQUENCE 313 AA; 33238 MW; 99D5C6E1C653941B CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 313;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATIAI 40
DB 83 DDATIAI 89
|||||

```

```

RESULT 50
O8XTH7 PRELIMINARY; PRT; 316 AA.
AC O8XTH7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Probable transmembrane protein.
GN RSP0133 OR RS02987.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Trehault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646076; CAD17284.1; -.
DR InterPro: IPR000620; DUF6.
DR InterPro: IPR000566; Lipocln_cytfabp.
DR Pfam: PF00892; DUF6; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KM Plasmid; Complete proteome.
SQ SEQUENCE 316 AA; 34411 MW; 0EF22D71759F791D CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 316;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LIAAVAF 12
DB 33 LIAAVAF 39
|||||

```

Search completed: February 21, 2003, 10:15:53
 Job time : 36 secs

```

RA Kantor A., Melt A., Shapira R.;
RT "Pedicoccus pentosaceus pediocin A encoding plasmid, PMD136.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF033858; AAD39627.1; -.
DR EMBL: AF063302; AAD25904.1; -.
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF000005; ABC_tran; 1.
DR Prodom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
KW ATP-binding; Plasmid.
SQ
SEQUENCE 292 AA; 32454 MW; 7A7F3B3EAE2B6047 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 SLEISDE 149
DB 23 SLEISDE 29

RESULT 45
O96169 PRELIMINARY; PRT; 300 AA.
AC O96169;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RNA-binding protein (KH domain).
GN PFB0370C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteil H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Petrea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL: AE001390; AAC71863.1; -.
DR InterPro: IPR004087; KH_dom.
DR SMART: SM00322; KH; 1.
SQ
SEQUENCE 300 AA; 36075 MW; 079F60A811CC3242 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 KRELEKN 213
DB 214 KRELEKN 220

RESULT 46
O9CB07 PRELIMINARY; PRT; 305 AA.
AC O9CB07;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative ABC transporter protein, ATP-binding component.
GN ML1726.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL: AL583923; CAC30679.1; -.
DR Leproma; ML1726; -.
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF000005; ABC_tran; 1.
DR Prodom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
KW ATP-binding; Complete proteome.
SQ
SEQUENCE 305 AA; 33197 MW; 865EBF8530D085E3 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 LSDVLT 188
DB 264 LSDVLT 270

RESULT 47
O9WMA2 PRELIMINARY; PRT; 307 AA.
AC O9WMA2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Ornithine cyclodeaminase (fragment).
GN OCD.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bacteria; Proteobacteria; Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=336;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K84;
RA Oger P.M., Farrand S.K.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF065244; AAD31597.1; -.
DR InterPro: IPR003462; ODC_Mu_crystall.
DR Pfam: PF02423; ODC_Mu_crystall; 1.
KW Plasmid.
FT NON_TER
SQ
SEQUENCE 307 AA; 33632 MW; 482A33066B8702A2 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 FGVLSDV 185
DB 50 FGVLSDV 56

RESULT 48
O8ZF01 PRELIMINARY; PRT; 310 AA.
AC O8ZF01;
DT 01-MAR-2002 (TREMBLrel. 20, Created)

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL: AE006715; AK41224.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 286 AA; 32278 MW; 71D00DBE5D978028 CRC64;

Query Match 3.3%; Score 7; DB 17; Length 286;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 206 FKRELEK 212
 |||||
 DB 259 FKRELEK 265

RESULT 41

O9R468 PRELIMINARY; PRT; 289 AA.
 AC O9R468;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Ornithine cyclodeaminase.
 GN TORP21.
 OS Agrobacterium tumefaciens.
 OG Plasmid pTIC58.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxId=358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Glenn J., Terryn N., Van Montagu M., Villarroel R.;
 RT "Complete nucleotide sequence of the T-DNA region of the plant tumor
 RT inducing Agrobacterium tumefaciens Ti plasmid pTIC58.";
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ237588; CAB4645.1; -
 DR InterPro: IPR002106; AtrRNA_LigaseII.
 DR InterPro: IPR003462; ODC_Mu_crystal1.
 DR Pfam: PF02423; ODC_Mu_crystal1.1.
 DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
 KW Plasmid.
 SQ SEQUENCE 289 AA; 31620 MW; 905A9B13022B9E5D CRC64;

Query Match 3.3%; Score 7; DB 2; Length 289;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 FGVLSPV 185
 |||||
 DB 31 FGVLSDV 37

RESULT 42

O8TVG6 PRELIMINARY; PRT; 289 AA.
 AC O8TVG6;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Uncharacterized protein.
 GN MK1423.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyracae;
 OC Methanopyrus.
 OX NCBI_TaxId=2320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozayavkin S.A.;
 RA "The complete genome of hyperthermophile Methanopyrus kandleri AV19

RT and monophyly of archaeal methanogens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL: AE010434; AAM02636.1; -
 KW Complete proteome.
 SQ SEQUENCE 289 AA; 30485 MW; 9F34178D2E6366FB CRC64;

Query Match 3.3%; Score 7; DB 17; Length 289;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 LSDVLTA 188
 |||||
 DB 95 LSDVLTA 101

RESULT 43

O8VLX7 PRELIMINARY; PRT; 291 AA.
 AC O8VLX7;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CzRB protein.
 GN CZRB.
 OS Thermus thermophilus.
 OC Bacteria; Thermus/Delnococcus group; Delnococci; Thermales;
 OC Thermaceae; Thermus.
 OX NCBI_TaxId=274;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Spada S., Pembroke J.T., Wall J.G.;
 RT "Cloning and characterisation of the czrb metal cation efflux protein
 RT from T. thermophilus.";
 RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ307316; CAC83722.1; -
 DR InterPro: IPR002524; Cation_efflux.
 DR Pfam: PF01545; Cation_efflux; 1.
 SQ SEQUENCE 291 AA; 31233 MW; 21CBA61D9DC8FB73 CRC64;

Query Match 3.3%; Score 7; DB 2; Length 291;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 VLSDVLT 187
 |||||
 DB 147 VLSDVLT 153

RESULT 44

O9MW33 PRELIMINARY; PRT; 292 AA.
 AC O9MW33;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative ABC transporter ATP-binding subunit.
 GN PENJ.
 OS Pedicoccus pentosaceus.
 OG Plasmid pMD136.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Lactobacillaceae; Pedicoccus.
 OX NCBI_TaxId=1255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC43200;
 RA Giacomini A., Marcazzan G., Salvato P., Squartini A., Nuti M.P.;
 RT "Nucleotide sequence of plasmid pMD136 from Pedicoccus pentosaceus
 RT ATCC43200.";
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC43200;

DR PRODOM: PD004758; Radical, activat. 1.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 KW Lyase.
 SQ SEQUENCE 270 AA; 29014 MW; 57F5C20CDBE2465 CRC64;
 Query Match 3.3%; Score 7; DB 2; Length 270;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 118 LAYKIGD 124
 DB 240 LAYKIGD 246
 RESULT 38
 098026 PRELIMINARY: PRT: 270 AA.
 AC 098026;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE MHC class I chain-related protein (Fragment).
 GN MHC.
 OS Hylobates lar (Common gibbon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
 OX NCBI_TaxID=9580;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99261547; PubMed=10331162;
 RA Pellet P., Vaneensbergh C., Debre P., Sumyuen M.H., Theodorou I.;
 RT "MHC genes in non-human primates.";
 RL Eur. J. Immunogenet. 26:239-241(1999).
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
 IMMUNE SYSTEM (BY SIMILARITY).
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
 MICROGLOBULIN) (BY SIMILARITY).
 CC EMBL: AF045604; AAD11614.1; -.
 DR InterPro: IPR003597; IG.CL.
 DR InterPro: IPR003006; IG.MHC.
 DR InterPro: IPR001039; MHC_I.
 DR Pfam: PF00047; Ig. 1.
 DR Pfam: PF00129; MHC_I. 1.
 DR PRINTS: PR01638; MHCCLASSI.
 DR PRODOM: PD000050; MHC_I. 1.
 DR SMART: SM00407; IGc1. 1.
 DR PROSITE: PS00290; IG.MHC; UNKNOWN_1.
 KW Glycoprotein; Transmembrane.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 270 AA; 31026 MW; AAD5DC05F1DBE5F CRC64;
 Query Match 3.3%; Score 7; DB 7; Length 270;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 149 EGNITMT 155
 DB 190 EGNITMT 196
 RESULT 39
 09D9W2 PRELIMINARY: PRT: 282 AA.
 AC 09D9W2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 1700026N20RIK protein.
 GN 1700026N20RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glisic C., King B., Kochia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guslincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nombarts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK006398; BAB24568.1; -.
 DR HSSP: P28867; IPR0.
 DR MGD: MGI:1917243; 1700026N20RIK.
 DR InterPro: IPR002219; DAG_PE-bind.
 DR InterPro: IPR000198; RhogAP.
 DR Pfam: PF00130; DAG_PE-bind. 1.
 DR Pfam: PF00620; RhogAP. 1.
 DR PRINTS: PR00008; DAGPEDOMAIN.
 DR SMART: SM00109; Cl. 1.
 DR SMART: SM00324; RhogAP. 1.
 DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
 SQ SEQUENCE 282 AA; 32540 MW; C53D6D40291FAEF0 CRC64;
 Query Match 3.3%; Score 7; DB 11; Length 282;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 76 IEARGLK 82
 DB 115 IEARGLK 121
 RESULT 40
 097ZG6 PRELIMINARY: PRT: 286 AA.
 AC 097ZG6;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE Hypothetical protein SS00950.
 GN SS00950.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Contafionieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Ehrano G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Kozeira C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";

Db 22 ISLEISD 28

RESULT 34

OBSZB4 PRELIMINARY: PRT: 252 AA.

DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE RE08075P.

GN CG13618.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Empidoidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celisner S.;
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY070987; AAL48609.1; -;
 SQ SEQUENCE 252 AA; 28845 MW; E7442C388865A5A CRC64;

Query Match 3.3%; Score 7; DB 5; Length 252;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LIIAIVA 11
 Db 11 LIIAIVA 17

RESULT 35

O9KSH1 PRELIMINARY: PRT: 252 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein VCI285.

GN VCI285.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unanue L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Yamathyan J., Bass S., Qin H., Dragol I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae";
 RL Nature 406:477-483(2000).
 DR EMBL: AE004207; AAF94444.1; -;
 DR TIGR: VCI285; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 252 AA; 28663 MW; 5E5E2C734D826B80 CRC64;

Query Match 3.3%; Score 7; DB 16; Length 252;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 GIVKAHL 103
 Db 19 GIVKAHL 25

RESULT 36

O9ER57 PRELIMINARY: PRT: 263 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Beta chimaerin.

GN 1700026N20R1K OR BCH.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Primer A., Heinlein U.A.O.;
 RT "Characterization of mouse beta chimaerin";
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ279014; CAC08453.1; -;
 DR HSSP: P2867; 1PTO.

DR MGD: MGI:1917243; 1700026N20R1K.
 DR InterPro: IPR002219; DAG_PE-bind.
 DR InterPro: IPR001198; RhogAP.
 DR Pfam: PF00130; DAG_PE-bind; 1.
 DR Pfam: PF00620; RhogAP; 1.
 DR PRINTS: PR00008; DAGPEDOMAIN.

DR SMART: SM00109; C1; 1.
 DR SMART: SM00324; RhogAP; 1.
 DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
 SQ SEQUENCE 263 AA; 30028 MW; 373D26ECD22971F CRC64;

Query Match 3.3%; Score 7; DB 11; Length 263;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 IEARGLK 82
 Db 110 IEARGLK 116

RESULT 37

O9RH16 PRELIMINARY: PRT: 270 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Pyruvate formate lyase activating enzyme.

GN PFLA.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-ZM4;
 RA Lee J.S., Kang H.S.;
 RT "Sequence analysis of 6563 cosmid clone of Zymomonas mobilis ZM4
 containing rna operon";
 RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF088897; AAF18276.1; -;
 DR InterPro: IPR000345; Cytochrome_bbind.
 DR InterPro: IPR000385; MoaA_NiFB_pgqE.
 DR InterPro: IPR001989; Radical_activat.

Pfam: PF01444; MoaA_NiFB_pgqE; 1.

RT genome.":
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF204951; AAK14454.1; -.
 SQ SEQUENCE 204 AA; 22852 MW; 03057C3A8015EE1 CRC64;

Query Match 3.3%; Score 7; DB 12; Length 204;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 107 VHDDIVS 113
 |||||
 Db 52 VHDDIVS 58

RESULT 30

O9UDJ0 PRELIMINARY: PRT: 214 AA.
 AC O9UDJ0:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE C380A1.1 (Novel protein) (Fragment).
 CN C380A1.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wallis J.;
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: 297653; CAB56187.1; -.
 FT NON-TER 1 214
 FT SEQUENCE 214 AA; 23405 MW; 269F5EC71887B96 CRC64;

Query Match 3.3%; Score 7; DB 4; Length 214;
 Best Local Similarity 100.0%; Pred. No. 91;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 DDAAIAI 40
 |||||
 Db 146 DDAAIAI 152

RESULT 31

O9LDP3 PRELIMINARY: PRT: 218 AA.
 AC O9LDP3:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE GblAAE32471.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-COLUMBIA;
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN-COLUMBIA;
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AB025605; BAA98058.1; -.

DR EMBL: AP002029; BAA97553.1; -.
 SQ SEQUENCE 218 AA; 24485 MW; 71C7AEBA109BDDA44 CRC64;

Query Match 3.3%; Score 7; DB 10; Length 218;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 VLTAFIQ 191
 |||||
 Db 120 VLTAFIQ 126

RESULT 32

O95DZ4 PRELIMINARY: PRT: 225 AA.
 AC O95DZ4:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Photosystem II CP47 protein (Fragment).
 OS Taxus chinensis var. mairei.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Taxaceae; Taxus.
 OX NCBI_TaxID=120273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lung J., Chang S.-H., Tsai C.-J., Ho C.-K.;
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY043261; AAK85713.1; -.
 DR Interpro: IPR000932; PSIIprot.
 DR Pfam: PF00421; PSII; 1.
 KW Chloroplast.
 FT NON-TER 1 1
 FT SEQUENCE 225 AA; 25060 MW; CD1CDC551E6789E8 CRC64;

Query Match 3.3%; Score 7; DB 8; Length 225;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 EEGIVKA 101
 |||||
 Db 90 EEGIVKA 96

RESULT 33

O9HEI0 PRELIMINARY: PRT: 250 AA.
 AC O9HEI0:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Putative guanine nucleotide binding protein similar to yor223w.
 GN SPAC20H4.02.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-972H-;
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL512487; CAC19732.1; -.
 SQ SEQUENCE 250 AA; 28262 MW; 1AD18FD0D28F3E63 CRC64;

Query Match 3.3%; Score 7; DB 3; Length 250;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 LSLEISD 148
 |||||

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC025724; AAC33373.1; -.
KW Hypothetical protein.
SQ SEQUENCE 192 AA; 20416 MW; 05FB571237A0A9D3 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ELLIAV 10
DB 10 ELLIAV 16

RESULT 26
Q97VF9 PRELIMINARY; PRT; 202 AA.
AC Q97VF9;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Hypothetical protein SS02667.
GN SS02667.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35092 / DSM 1617 / P2;
RC MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awey M.J., Chan-Weiler C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006861; AAK42785.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 202 AA; 23604 MW; D5298B190984AE5A CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YDLAYKL 122
DB 141 YDLAYKL 147

RESULT 27
Q99PT1 PRELIMINARY; PRT; 204 AA.
AC Q99PT1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RhogDI-1.
GN C87222 OR GDI-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Minamitani T., Matsumoto K.;
RT "Mouse cDNA sequence for RhogDI-1."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055070; BAB21527.1; -.
DR HSSP; P19803; 1GDF.
DR PMMA-2DPAGE; Q99PT1; -.
DR MGD; MGI:2144689; C87222.
DR InterPro; IPR000406; Rho_GDI.
DR Pfam; PF02115; Rho_GDI; 1.
DR PRINTS; PR00492; RHOGDI.
SQ SEQUENCE 204 AA; 23407 MW; 8ACE6F4456D842D8 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
DB 59 VAVSADP 65

RESULT 28
Q99KC4 PRELIMINARY; PRT; 204 AA.
AC Q99KC4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 23.4 kDa protein.
GN C87222.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strussberg R.;
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004732; AAH04732.1; -.
DR HSSP; P19803; 1GDF.
DR MGD; MGI:2144689; C87222.
DR InterPro; IPR000406; Rho_GDI.
DR Pfam; PF02115; Rho_GDI; 1.
DR PRINTS; PR00492; RHOGDI.
KW Hypothetical protein.
SQ SEQUENCE 204 AA; 23391 MW; 8B126F59970D8250 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
DB 59 VAVSADP 65

RESULT 29
Q8QNN3 PRELIMINARY; PRT; 204 AA.
AC Q8QNN3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Esv-1-28.
OS Ectocarpus siliculosus virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.
OX NCBI_TaxID=37665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Esv-1;
RA Delaroque N., Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W.;
RT "The complete nucleotide sequence of the Ectocarpus siliculosus virus

```

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
 RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL391148; CAC01866.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 183 AA; 19958 MW; 88367943EC353AD4 CRC64;
 QY 185 VLTAFQ 191
 Db 87 VLTAFQ 93
 RESULT 23
 ID 09M878 PRELIMINARY; PRT; 185 AA.
 AC 09M878;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Pf683.27 protein (Hypothetical 20.4 kDa protein).
 GN Pf683.27.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
 RA Rongning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III BAC F1683 genomic sequence.",
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carlincl P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kaniya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shin P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC021640; AAP32471.1; -
 DR EMBL: AY062624; AAP32702.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 185 AA; 20382 MW; F3020E240D8C02E6 CRC64;
 QY 185 VLTAFQ 191
 Db 88 VLTAFQ 94
 Query Match 3.3%; Score 7; DB 10; Length 185;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 24
 ID 08U8T1 PRELIMINARY; PRT; 190 AA.
 AC 08U8T1;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Ornithine cyclodeaminase (N-terminal).
 DE

GN ARCB* OR ATU4008 OR AGR_L1691.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 RX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58.",
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlet K., Gordon B., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Woliam C., Allinger M., Doughty D., Scott C., Iappas C., Markelz B.,
 RA Planagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.",
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009331; AAL44809.1; -
 DR EMBL: AE008283; AAK89418.1; -
 KW Complete proteome.
 SQ SEQUENCE 190 AA; 20766 MW; A1D1D6A86D0AF0FB CRC64;
 QY 179 FGVLSDV 185
 Db 96 FGVLSDV 102
 Query Match 3.3%; Score 7; DB 16; Length 190;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 25
 ID 09GUN5 PRELIMINARY; PRT; 192 AA.
 AC 09GUN5;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Hypothetical 20.4 kDa protein.
 GN Y67D8C.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 RX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.",
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Edwards J., Du H., Lamar B., Wohldmann P., Walker C.;
 RT "The sequence of C. elegans cosmid Y67D8C.",

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFP303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003007; BAB52281.1; -
 DR InterPro: IPR004952; DUF269.
 DR Pfam: PF03270; DUF269; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 159 AA; 17409 MW; 3A1B615047570878 CRC64;

Query Match 3.3%; Score 7; DB 16; Length 159;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDAAIAI 40
 Db 144 DDAAIAI 150

RESULT 19
 Q47896
 ID 047896 PRELIMINARY; PRT; 169 AA.
 AC 047896;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE ORF 169.
 OS Tolypotrrix sp. PCC 7601.
 OC Bacteria; Cyanobacteria; Nostocales; Microchaetaceae; Tolypotrrix.
 OX NCBI_TaxID=1188;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88260883; PubMed=2838727;
 RA Parsot C., Mazel D.;
 RT "Cloning and nucleotide sequence of the thrB gene from the
 RT cyanobacterium Calothrix PCC 7601.";
 RL MBL: Y00522; CAA68577.1; -
 DR EMBL: Y00522; CAA68577.1; -
 SQ SEQUENCE 169 AA; 19182 MW; 1A036CFA6CEAA00 CRC64;

Query Match 3.3%; Score 7; DB 2; Length 169;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 IDDAIAA 39
 Db 62 IDDAIAA 68

RESULT 20
 Q8RM78
 ID 08RM78 PRELIMINARY; PRT; 176 AA.
 AC 08RM78;
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Hypothetical 20.5 kDa protein.
 OS Bacteroides fragilis.
 OC Bacteri; CFB group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
 OX Bacterioides.
 OX NCBI_TaxID=817;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-LV25; TRANSPOSON-CLV25;
 RX MEDLINE=21886381; PubMed=11889096;
 RA Bass K.A., Hecht D.W.;
 RT "Isolation and Characterization of CLV25, a Bacterioides fragilis
 RT Chromosomal Transfer Factor Resembling Multiple Bacterioides sp.
 RT Mobilizable Transposons.";
 RL J. Bacteriol. 184:1895-1904(2002).
 DR EMBL: AY053505; AAL29897.1; -
 KW Hypothetical protein
 SQ SEQUENCE 176 AA; 20541 MW; C608B6C209FDE10 CRC64;

Query Match 3.3%; Score 7; DB 2; Length 176;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 AFKRELE 211
 Db 13 AFKRELE 19

RESULT 21
 Q9A078
 ID 09A078 PRELIMINARY; PRT; 178 AA.
 AC 09A078;
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE Hypothetical protein SPY0899.
 GN SPY0899.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Szatze S., Suvorov A.N., Kenyon S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL: AE006539; AAK33816.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 178 AA; 20463 MW; 86D29CEAC4B9703A CRC64;

Query Match 3.3%; Score 7; DB 16; Length 178;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 FGVLSDV 185
 Db 123 FGVLSDV 129

RESULT 22
 Q9LF03
 ID 09LF03 PRELIMINARY; PRT; 183 AA.
 AC 09LF03;
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
 DE Hypothetical 20.0 kDa protein.
 GN T21H19_170.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.

RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
 RL DNA Ref. 5:55-76(1998).
 DR EMBL: AP000007; BAA30975.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 140 AA; 16226 MW; CA70FD32DB80860A CRC64;

Query Match 3.3%; Score 7; DB 17; Length 140;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 TREBINK 31
 DB 8 TREBINK 14

RESULT 15

OTRX30 PRELIMINARY; PRT; 142 AA.

AC OTRX30;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
 DE Membrane protein implicated in regulation of membrane protease activity.
 GN MK0848.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 OC Methanopyrus
 OC NCBI_Taxid=2320;

RM [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
 RA Shecterlinina O.V., Shakova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.B., Tatsovsk R.L., Wolf Y.I., Stetter K.O.,
 RA Matlykh A.G., Koonin E.V., Kozyskin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 RT and monophyly of archaeal methanogens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL: AE010374; AAM02061.1; -.
 KW Protease; Complete proteome.
 SQ SEQUENCE 142 AA; 15533 MW; 34AE784117796477 CRC64;

Query Match 3.3%; Score 7; DB 17; Length 142;

Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 KEGEIV 99
 DB 86 KEGEIV 92

RESULT 16

O9HPE8 PRELIMINARY; PRT; 151 AA.

AC O9HPE8;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DE Vng1670c.
 GN VNG1670C.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OC NCBI_Taxid=64091;

RM [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahatras G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
 RA Lelthausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,

RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angwine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Eshardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarra S.;
 RT "Genome sequence of Halobacterium species NRC-1."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL: AE005076; AAG19921.1; -.
 KW Complete proteome.

SQ SEQUENCE 151 AA; 15766 MW; 87585DA4F36AC5AA CRC64;

Query Match 3.3%; Score 7; DB 17; Length 151;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 138 FVALSL 144
 DB 24 FVALSL 30

RESULT 17

O8TLH7 PRELIMINARY; PRT; 157 AA.

AC O8TLH7;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
 DE Hypothetical protein MA3059.
 GN MA3059.

OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OC NCBI_Taxid=2214;

RM [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbach C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann N., Deatellano K., Johnson R.,
 RA Linton L., McSwan P., McKernan K., Talamas J., Tirelli A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL: AE011008; AAM06432.1; -.
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 157 AA; 17167 MW; 32C8A097FA6C1602 CRC64;

Query Match 3.3%; Score 7; DB 17; Length 157;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 IAAVAV 13
 DB 17 IAAVAV 23

RESULT 18

O98AP1 PRELIMINARY; PRT; 159 AA.

AC O98AP1;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)
 DE Hypothetical protein mlr5912.
 GN MLR5912.
 OS Rhizobium loti (Mesorhizobium loti).

RN [1]
 RP SEQUENCE FROM N.A.
 RA van den Bogaard P.T.C., van Wieringen M., Memelink K., Kleerebezem M.,
 RA Kuipers O.P., de Vos W.M.;
 RT "Characterization of the Streptococcus thermophilus pTSH operon and
 RT analysis of HPr mutants.";
 RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY064171; AAL47557.1;
 DR InterPro: IPR001020; HPr_HisP_site.
 DR InterPro: IPR000032; HPr_protein.
 DR InterPro: IPR002114; HPr_Serp_site.
 DR Pfam: PF00381; PTS-HPr; 1.
 DR PRINTS: PR00107; PHOSPHOCPPR.
 DR PROSITE: PS00369; PTS_HPR_HIS; UNKNOWN_1.
 DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
 SQ SEQUENCE 87 AA; 8919 MW; 77DAF28C2E9F08DE CRC64;

Query Match 3.3%; Score 7; DB 2; Length 87;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 DDAIAI 40
 DB 71 DDAIAI 77

RESULT 12
 O99265 PRELIMINARY; PRT; 87 AA.
 AC O99265;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative phosphotransferase system phosphohistidine-containing
 DE protein.
 GN PTH OR SPY1373.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE-21192684; PubMed-11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najat F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL: AE006575; AAK34200.1;
 DR HSSP: P07515; IPTF.
 DR InterPro: IPR001020; HPr_HisP_site.
 DR InterPro: IPR000032; HPr_protein.
 DR InterPro: IPR002114; HPr_Serp_site.
 DR Pfam: PF00381; PTS-HPr; 1.
 DR PRINTS: PR00107; PHOSPHOCPPR.
 DR TIGRfams: TIGR01003; PTS_HPR_family; 1.
 DR PROSITE: PS00369; PTS_HPR_HIS; 1.
 DR PROSITE: PS00589; PTS_HPR_SER; 1.
 KW Transferrase; Complete proteome.
 SQ SEQUENCE 87 AA; 8951 MW; 760A961843464AE CRC64;

Query Match 3.3%; Score 7; DB 16; Length 87;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 DDAIAI 41
 DB 72 DDAIAI 78

RESULT 13

O970M6
 ID O970M6 PRELIMINARY; PRT; 87 AA.
 AC O970M6;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Phosphocarrier protein HPr.
 GN SP1177.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TIGR4;
 RX MEDLINE-21357209; PubMed-11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouli H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae.";
 RL Science 293:498-506(2001).
 DR EMBL: AE007418; AAK75286.1;
 DR TIGR: SP1177;
 DR InterPro: IPR001020; HPr_HisP_site.
 DR InterPro: IPR000032; HPr_protein.
 DR InterPro: IPR002114; HPr_Serp_site.
 DR Pfam: PF00381; PTS-HPr; 1.
 DR PRINTS: PR00107; PHOSPHOCPPR.
 DR TIGRfams: TIGR01003; PTS_HPR_family; 1.
 DR PROSITE: PS00369; PTS_HPR_HIS; 1.
 DR PROSITE: PS00589; PTS_HPR_SER; 1.
 KW Complete proteome.
 SQ SEQUENCE 87 AA; 8939 MW; 2D610EAEAF25AF70 CRC64;

Query Match 3.3%; Score 7; DB 16; Length 87;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 DDAIAI 40
 DB 71 DDAIAI 77

RESULT 14
 O59502 PRELIMINARY; PRT; 140 AA.
 AC O59502;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein PH1854.
 GN PH1854.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE-98344137; PubMed-9679194;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Hailkawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuku Y.,
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-


```

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Gag-pol polyprotein gp80 precursor.
OS Feline leukemia virus.
OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus.
NCBI_TaxId=11768;

RN
RP SEQUENCE FROM N.A.
RC STRAIN=FEV-FAIDS;
MEDLINE=88119207; PubMed=2828667;
RA Donahue P.R., Hoover E.A., Belz G.A., Riedel N., Hirsch V.M.,
RA Overbaugh J., Mullins J.I.;
RT "Strong sequence conservation among horizontally transmissible,
RT minimally pathogenic feline leukemia viruses."
RL J. Virol. 62:722-731(1988).
CC -I SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC KNOWN AS THE RETROPEPSIN FAMILY.

DR EMBL; M18247; AAA93092.1; -.
DR HSSP; P03355; 1MML.
DR InterPro: IPR001995; Aspprotease_rtrv.
DR InterPro: IPR001969; Aspprotease_site.
DR InterPro: IPR000840; Gag_MA.
DR InterPro: IPR002079; Gag_P12.
DR InterPro: IPR003036; Gag_P30.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam; PF01140; Gag_MA; 1.
DR Pfam; PF01141; Gag_P12; 1.
DR Pfam; PF02093; Gag_P30; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00098; Zf_CCHC; 1.
DR SMART; SM00343; ZnF_C2HC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS50175; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Core protein; Hydrolyase; Polyprotein;
KW RNA-directed DNA polymerase; Signal.
FT SIGNAL 1766 1784
FT CHAIN 75 576
FT CHAIN 75 201
FT CHAIN 202 271
FT CHAIN 272 519
FT CHAIN 520 576
FT CHAIN 577 1784
FT CHAIN 577 701
FT CHAIN 702 1368
FT CHAIN 1369 1784
SQ SEQUENCE 1784 AA; 200157 MW; D8162E567D054688 CRC64;

Query Match 3.8%; Score 8; DB 15; Length 1784;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KEMTKVLA 203
DB 513 KEMTKVLA 520

```

```

RN
RP [1]
RC SEQUENCE FROM N.A.
RC STRAIN=RICKARD SUBGROUP A;
RX MEDLINE=98362106; PubMed=9696797;
RA Chen H., Bechtel M.K., Shi Y., Phlips A., Mathes L.E., Hayes K.A.,
RA Roy-Burman P.;
RT "Pathogenicity induced by feline leukemia virus, rickard strain,
RT subgroup A plasmid DNA (pFRA).";
RL J. Virol. 72:7048-7056(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=RICKARD SUBGROUP A;
RA Chen H., Roy-Burman P.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
CC -I SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC KNOWN AS THE RETROPEPSIN FAMILY.

DR EMBL; AF052723; AAC31801.1; -.
DR HSSP; P03355; 1MML.
DR InterPro: IPR001995; Aspprotease_rtrv.
DR InterPro: IPR001969; Aspprotease_site.
DR InterPro: IPR000840; Gag_MA.
DR InterPro: IPR002079; Gag_P12.
DR InterPro: IPR003036; Gag_P30.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam; PF01140; Gag_MA; 1.
DR Pfam; PF01141; Gag_P12; 1.
DR Pfam; PF02093; Gag_P30; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00098; Zf_CCHC; 1.
DR SMART; SM00343; ZnF_C2HC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS50175; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Core protein; Hydrolyase; Polyprotein;
KW RNA-directed DNA polymerase.
FT CHAIN 75 576
FT CHAIN 75 201
FT CHAIN 202 271
FT CHAIN 272 519
FT CHAIN 520 576
FT CHAIN 577 1786
FT CHAIN 577 701
FT CHAIN 702 1370
FT CHAIN 1371 1786
SQ SEQUENCE 1786 AA; 200201 MW; 49C46E78722DEF18 CRC64;

Query Match 3.8%; Score 8; DB 15; Length 1786;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KEMTKVLA 203
DB 513 KEMTKVLA 520

```

```

RESULT 10
ID 089811 PRELIMINARY; PRT; 1786 AA.
AC 089811;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Gag-pol polyprotein gp80.
OS Feline leukemia virus.
OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus.
NCBI_TaxId=11768;

```

```

RESULT 11
ID 08VP89 PRELIMINARY; PRT; 87 AA.
AC 08VP89;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE HPr.
GN PTH.
OS Streptococcus thermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
NCBI_TaxId=1308;

```

RL Gene 122:63-70(1992).
 CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
 DR EMBL: M90412; AAA26764.1; -.
 DR HSSP: P00579; 1SIG.
 DR InterPro: IPR000901; CPase.
 DR InterPro: IPR00943; Sigma_70.
 DR Pfam: PF00140; sigma70.1.
 DR PRINTS: PR00046; SIGMA70ECT.
 DR PROSITE: PS00867; CPASE_2; UNKNOWN_1.
 DR PROSITE: PS00715; SIGMA70.1; 1.
 DR PROSITE: PS00716; SIGMA70.2; 1.
 KW DNA-binding; DNA-directed RNA polymerase; Sigma factor;
 KW Transcription regulation.
 SQ SEQUENCE 528 AA; 57598 MW; A22B76FE5C065D4 CRC64;

Query Match 3.8%; Score 8; DB 2; Length 528;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 204 PAFKRELE 211
 Db 275 PAFKRELE 282

RESULT 6
 Q93Q20 PRELIMINARY; PRT; 632 AA.
 AC Q93Q20;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Pknc.
 OS Nostoc punctiforme.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=63737;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wong F.C.Y., Weeks J.C.;
 RT "A novel gene required for normal heterocyst differentiation pattern
 RT in Nostoc punctiforme.";
 RT Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF288131; AAK68687.1; -.
 DR InterPro: IPR00719; Euk_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR PRODOM: PD000001; Euk_pkinase; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Transferase.
 SQ SEQUENCE 632 AA; 69868 MW; C1F4321AB4970E87 CRC64;

Query Match 3.8%; Score 8; DB 2; Length 632;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 EINKATDD 35
 Db 545 EINKATDD 552

RESULT 7
 Q8S8T2 PRELIMINARY; PRT; 774 AA.
 AC Q8S8T2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Subtilisin-like serine protease.
 GN ATG39850.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Somerville C.R., Venter J.C.;
 RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC003000; AAM14853.1; -.
 KW Protease.
 SQ SEQUENCE 774 AA; 87729 MW; 7F570AA2FAFD8DD2 CRC64;

Query Match 3.8%; Score 8; DB 10; Length 774;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 KAIDDAIA 38
 Db 257 KAIDDAIA 264

RESULT 8
 O24835 PRELIMINARY; PRT; 918 AA.
 AC O24835;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 93.4 kDa protein (Fragment).
 OS Acinetobacter sp. ADP1.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 OC Acinetobacter.
 OX NCBI_TaxID=62977;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE-99287833; PubMed-10348863;
 RA Segura A., Bunz P.V., D'Argenio D.A., Ornstom L.N.;
 RT "Genetic analysis of a chromosomal region containing vanA and vanB,
 RT genes required for conversion of either ferulate or vanillate to
 RT protocatechuate in Acinetobacter.";
 RT J. Bacteriol. 181:3494-3504(1999).
 DR EMBL: AF011339; AAC27114.1; -.
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 1
 FT NON_TER 918
 FT NON_TER 918
 SQ SEQUENCE 918 AA; 93381 MW; 2991F95E5B7DB94E CRC64;

Query Match 3.8%; Score 8; DB 2; Length 918;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 INKAIDA 36
 Db 78 INKAIDA 85

RESULT 9
 Q85521 PRELIMINARY; PRT; 1784 AA.
 AC Q85521;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

QY 188 AIFODTVRK 196
 DB 86 AIFODTVRK 94

RESULT 2

Q85514 PRELIMINARY: PRT; 68 AA.

AC Q85514;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 OS (H2S-FesV) recombinational site DNA (Fragment).
 OS Feline sarcoma virus.
 OS Viruses; Retrovirus; Retroviridae; Mammalian type C retroviruses.
 OC NCBI_TaxID=11772;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86308237; PubMed=3018286;
 RA Besmer P., Lader E., George P.C., Bergold P.J., Qiu F.-H.,
 RA Zuckerman E.E., Hardy W.D.;
 RT "A new acute transforming feline retrovirus with fms homology
 RT specifies a C-terminally truncated version of the c-fms protein that
 RT is different from SM-Feline sarcoma virus v-fms protein."
 RL J. Virol. 60:194-203(1986).
 DR EMBL; M14290; AAA43043.1; .
 FT NON_TER 1 1
 FT SEQUENCE 68 AA; 7141 MW; 1EF902F883F9ECBC CRC64;

Query Match

Best Local Similarity 3.8%; Score 8; DB 15; Length 68;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KEMTKVLA 203
 DB 2 KEMTKVLA 9

RESULT 3

Q9S066 PRELIMINARY: PRT; 349 AA.

AC Q9S066;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 38.7 kDa protein.
 GN T17F15.100.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choisne N., Robert C., Brotier P., Wincker P., Cattolico L.,
 RA Artiguenave F., Saurin W., Weissbach J., Mewes H.W., Meyer K.F.X.,
 RA Lemke K., Schueller C., Queller F., Salanoubat M.;
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 DR EMBL; AL049658; CAB41136.1; .
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00097; zf-C3HC4; 1.
 DR SMART; SM00184; RING; 1.
 KW Hypothetical protein; zinc-finger.
 SQ SEQUENCE 349 AA; 38686 MW; A080BD4CD9DEBBE CRC64;

Query Match 3.8%; Score 8; DB 10; Length 349;
 Best Local Similarity 100.0%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 136 QDFVALS 143
 DB 332 QDFVALS 339

RESULT 4

Q9RAM9 PRELIMINARY: PRT; 437 AA.

AC Q9RAM9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Homoserine dehydrogenase (EC 1.1.1.3) (HDH).
 GN HDH.
 OS Methylobacillus flagellatum.
 OC Bacteria; Proteobacteria; Beta subdivision; Methylobacillus group;
 OC Methylobacillus.
 OC NCBI_TaxID=405;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KT;
 RX MEDLINE=20055654; PubMed=10589737;
 RA Marchenko G.N., Marchenko N.D., Tsygankov Y.D., Chistoserlov A.Y.;
 RT "Organization of threonine biosynthesis genes from the obligate
 RT methylotriph Methylobacillus flagellatus."
 RL Microbiology 145:3273-3282(1999)
 CC -1- CATALYTIC ACTIVITY: L-HOMOSERINE + NAD(P)(+) = L-ASPARTATE 4-
 CC -1- SEMIALDEHYDE + NAD(P)H.
 CC -1- SIMILARITY: BELONGS TO THE HOMOSERINE DEHYDROGENASE FAMILY.
 DR EMBL; L78665; AAF21129.1; .
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR002202; HMG-COA_red.
 DR InterPro; IPR001342; Homoserine_dh.
 DR InterPro; IPR005106; NAD_binding_3.
 DR Pfam; PF01842; ACT; 1.
 DR Pfam; PF00742; Homoserine_dh; 1.
 DR Pfam; PF03447; NAD_binding_3; 1.
 DR PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
 DR PROSITE; PS01042; HOMOSER_DHEGENASE; 1.
 KW NADP; Oxidoreductase.
 SQ SEQUENCE 437 AA; 46702 MW; B130BFF465F9BB2 CRC64;

Query Match

Best Local Similarity 3.8%; Score 8; DB 2; Length 437;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDAAIAIE 41
 DB 411 DDAAIAIE 418

RESULT 5

Q59814 PRELIMINARY: PRT; 528 AA.

AC Q59814;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE RNA polymerase sigma factor.
 GN HRDC.
 OS Streptomyces aureofaciens.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OC NCBI_TaxID=1894;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CCM 3239;
 RX MEDLINE=93083996; PubMed=1452038;
 RA Kornanev J., Farkasovsky M., Potuckova L.;
 RT "Four genes in Streptomyces aureofaciens containing a domain
 RT characteristic of principal sigma factors.";

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:12:35 ; Search time 31 Seconds
(without alignments)
1415.743 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213

Sequence: 1 MKKFLIAVAFAVAVSADPI.....VRKEMTKVLAPFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	4.2	374	6 Q9TU81	Q9TU81 ovine aries
2	8	3.8	68	15 Q85514	Q85514 feline sarc
3	8	3.8	349	10 Q9SU06	Q9SU06 arabidopsis
4	8	3.8	437	2 Q9RAM9	Q9RAM9 methylobact
5	8	3.8	528	2 Q59814	Q59814 streptomyces
6	8	3.8	632	2 Q93020	Q93020 nostoc punc
7	8	3.8	774	10 Q858T2	Q858T2 arabidopsis
8	8	3.8	2	024835	024835 acinetobact
9	8	3.8	1784	15 Q85521	Q85521 feline leuk
10	8	3.8	1786	15 Q85811	Q85811 feline leuk
11	7	3.3	87	2 Q8VP89	Q8VP89 streptococc
12	7	3.3	87	16 Q99Z65	Q99Z65 streptococc
13	7	3.3	87	16 Q97QM6	Q97QM6 streptococc
14	7	3.3	140	17 Q59502	Q59502 pyrococcus
15	7	3.3	142	17 Q8TX30	Q8TX30 methanopyru
16	7	3.3	151	17 Q9HPE8	Q9HPE8 halobacteri

17	7	3.3	157	17 Q8TLH7	Q8TLH7 methanosarc
18	7	3.3	159	16 Q98AP1	Q98AP1 rhizobium 1
19	7	3.3	169	2 Q47896	Q47896 tolypotrixi
20	7	3.3	176	2 Q8RM78	Q8RM78 bacteroides
21	7	3.3	178	16 Q9A078	Q9A078 streptococc
22	7	3.3	183	10 Q9LF03	Q9LF03 arabidopsis
23	7	3.3	185	10 Q9M878	Q9M878 arabidopsis
24	7	3.3	190	16 Q8U8T1	Q8U8T1 agrobacteri
25	7	3.3	192	5 Q9GUNS	Q9GUNS caenorhabdi
26	7	3.3	202	17 Q97VF9	Q97VF9 sulfobolus
27	7	3.3	204	11 Q99PT1	Q99PT1 mus musculu
28	7	3.3	204	11 Q99KC4	Q99KC4 mus musculu
29	7	3.3	204	12 Q8QNN3	Q8QNN3 ectocarpus
30	7	3.3	214	4 Q9UJ10	Q9UJ10 homo sapien
31	7	3.3	218	10 Q9LDP3	Q9LDP3 arabidopsis
32	7	3.3	225	8 Q95D24	Q95D24 taxus chine
33	7	3.3	250	3 Q9HE10	Q9HE10 schizosacch
34	7	3.3	252	5 Q85ZB4	Q85ZB4 dirosophila
35	7	3.3	252	16 Q9KSH1	Q9KSH1 vibrio chol
36	7	3.3	263	11 Q9ER57	Q9ER57 mus musculu
37	7	3.3	270	2 Q9RH16	Q9RH16 zymomonas m
38	7	3.3	270	7 Q98026	Q98026 hylobates 1
39	7	3.3	282	11 Q9D9W2	Q9D9W2 mus musculu
40	7	3.3	286	17 Q97ZG6	Q97ZG6 sulfolobus
41	7	3.3	289	2 Q9R468	Q9R468 agrobacteri
42	7	3.3	289	17 Q8TVG6	Q8TVG6 methanopyru
43	7	3.3	291	2 Q8VLX7	Q8VLX7 thermus the
44	7	3.3	292	2 Q9WW33	Q9WW33 pediococcus
45	7	3.3	300	5 Q96169	Q96169 plasmodium

ALIGNMENTS

RESULT 1	ALIGNMENTS
Q9TU81	PRELIMINARY; PRT; 374 AA.
ID Q9TU81	
AC Q9TU81;	
DT 01-MAY-2000 (TREMBLrel. 13, Created)	
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE Pulmonary surfactant-associated protein B (Pfragment).	
OS Ovis aries (Sheep).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC Bovidae; Caprinae; Ovis.	
OX NCBI_TaxID=9940;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=LUNG;	
RX MEDLINE=20215263; PubMed=10749754;	
RA Pietschmann S.M., Plson U.;	
RT "CDNA Cloning of ovine pulmonary surfactant proteins A, B, and C:	
RT Isolation of two different sequences for SP-B";	
RT Am. J. Physiol. 278:L765-L778(2000).	
DR EMBL: AF107544; AAF14195.1; -	
DR HSSP: P07988; 1DFW.	
DR InterPro: IPR003119; SAPA.	
DR InterPro: IPR000004; SAPB.	
DR InterPro: IPR003258; Surfactant_B.	
DR Pfam: PF02199; SAPA; 1.	
DR Pfam: PF03489; Surfactant_B; 1.	
DR ProDom: PD001732; SAPB; 1.	
DR ProDom: PD008002; Surfactant_B; 1.	
DR SMART: SM00162; SAPA; 1.	
DR SMART: SM00118; SAPB; 3.	
FT NON_TER	
FT SEQUENCE	
SO	
Query Match	4.2%; Score 9; DB 6; Length 374;
Best Local Similarity	100.0%; Pred. No. 1.3;
Matches	9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: ACTIVATED T CELLS, MAST CELLS, NATURAL KILLER
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE IL-3 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M14744; AAA35455.1; -.
CC PIR: B24427; B24427.
CC HSSP: P08700; 1JLT.
CC InterPro: IPR002183; Interleukin_3.
CC Pfam: PF02059; IL3; 1.
CC PRINTS: PR00430; INTERLEUKIN3.
CC CYTOKINE; Glycoprotein; Growth factor; Signal.
KW SIGNAL
FT CHAIN 1 19 BY SIMILARITY.
FT DISULFD 20 152 INTERLEUKIN-3.
FT DISULFD 35 103 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 152 AA: 17207 MW: F63438A0D2577B3A CRC64;

Query Match 2.8%; Score 6; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 142 LSLEIS 147
   |||||
Db 147 LSLEIS 152

```

Search completed: February 21, 2003, 10:15:15
 Job time : 17 secs

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67471; AAB98125.1; -.
DR TIGR: MJ0142; -.
DR InterPro: IPR002806; DUF103.
DR Pfam: PF01953; DUF103; 1.
DR ProDom: PD015886; DUF103; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 147 AA; 17592 MW; 707D67BF058E065C CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 KRELEK 212
DB 11 KRELEK 16
|||||

RESULT 48
DUT_THETN STANDARD; PRT; 148 AA.
AC O8RA46;
DT 15-JUN-2002 (Rel. 41, Last Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23)
DE (dUTPase) (dUTP pyrophosphatase).
GN DUT OR TTEJ384.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RA MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RA "A complete sequence of T. tengcongensis genome.";
RT Genome Res. 12:689-700(2002).
CC -1- FUNCTION: This enzyme is involved in nucleotide metabolism: it
CC produces dUMP, the immediate precursor of thymidine nucleotides
CC and it decreases the intracellular concentration of dUTP so that
CC uracil cannot be incorporated into DNA (By similarity).
CC -1- CATALYTIC ACTIVITY: dUMP + H(2)O = dUMP + diphosphate.
CC -1- PATHWAY: De novo synthesis of thymidylate.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE013098; AAM24606.1; -.
KM Hydrolyase; Nucleotide metabolism; Complete proteome.
SQ SEQUENCE 148 AA; 16297 MW; A66B21AC647755AA CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 ANVKG 95
DB 33 ANVKG 38
|||||

```

```

RESULT 49
YMBW_YEAST STANDARD; PRT; 150 AA.
ID YMBW_YEAST
AC 003579;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 17.2 kDa protein in PRCl-ABD4 intergenic region.
GN YMR298W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X80836; CAA56807.1; -.
DR SGD: S0004913; YMR298W.
KM Hypothetical protein.
SQ SEQUENCE 150 AA; 17207 MW; 75DBA35225C3065C CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LITAAV 10
DB 32 LITAAV 37
|||||

RESULT 50
IL3_HYLLA STANDARD; PRT; 152 AA.
ID IL3_HYLLA
AC P06740;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Interleukin-3 precursor (IL-3) (Multipotential colony-stimulating
DE factor) (Hematopoietic growth factor) (P-cell stimulating factor)
DE (Mast-cell growth factor) (MCGF).
GN IL3.
OS Hylobates lar (Common gibbon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
OX NCBI_TaxID=9580;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=87002490; PubMed=3489530;
RA Yang Y.-C., Charietta A.B., Temple P.A., Chung M.P., Kovacic S.,
RA Witk-Gianotti J.S., Leary A.C., Kriz R., Donahue R.E., Wong G.G.,
RA Clark S.C.;
RT "Human IL-3 (mult-CSF): identification by expression cloning of a
RT novel hematopoietic growth factor related to murine IL-3.";
RL Cell 47:3-10(1986).
CC -1- FUNCTION: GRANULOCYTE/MACROPHAGE COLONY-STIMULATING FACTORS ARE
CC CYTOKINES THAT ACT IN HEMATOPOIESIS BY CONTROLLING THE PRODUCTION,
CC DIFFERENTIATION, AND FUNCTION OF 2 RELATED WHITE CELL POPULATIONS
CC OF THE BLOOD, THE GRANULOCYTES AND THE MONOCYTES-MACROPHAGES.
CC -1- FUNCTION: THIS CSF INDUCES GRANULOCYTES, MACROPHAGES, MAST CELLS,
CC STEM CELLS, ERYTHROID CELLS, EOSINOPHILS AND MEGAKARYOCYTES.

```

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Flischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jatali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Mlshina N.V., Moberly C., Morris J., Mosnrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palzert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissendach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- TISSUE SPECIFICITY: SOMATIC AND GERM LINE CELLS.
 CC -1- DEVELOPMENTAL STAGE: THE NON-SEX-SPECIFIC TRANSCRIPT IS PRESENT AT
 ALL STAGES.
 CC -1- MISCELLANEOUS: TRANSCRIPTION OF JANA GIVES RISE TO TWO
 CC -1- SEX-SPECIFIC AND TO ONE NON-SEX-SPECIFIC TRANSCRIPTS.
 CC -1- SIMILARITY: WITH SEX-REGULATED PROTEIN JANUS-B.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC CC
 DR EMBL: M27033; AAC34203.1; -
 DR EMBL: M27033; AAC34202.1; ALT_INIT.
 DR EMBL: AE003772; AAF56997.1; -
 DR PIR: A32317; A32317.
 DR FlyBase: FBgn0001280; jana.
 SQ SEQUENCE 135 AA; 15220 MW; 2720237CE77F3132 CRC64;
 QY 145 EISDEG 150
 DB 129 EISDEG 134
 Query Match 2.8%; Score 6; DB 1; Length 135;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OC Callitichthyidae; Hoplosternum.
 OX NCBI_TaxID=114109;
 RN [1]
 RP SEQUENCE, SUBUNIT, AND MASS SPECTROMETRY.
 RC TISSUE=Blood;
 RX MEDLINE=20298792; PubMed=10747999;
 RA Weber R.E., Fago A., Val A.L., Bang A., Van Hauwaert M.-L.,
 RA Dewilde S., Zal F., Moens L.;
 RT "Ischemoglobin differentiation in the bimodal-breathing amazon
 RT catfish *Hoplosternum littorale*.";
 RL J. Biol. Chem. 275:17297-17305(2000).
 CC -1- FUNCTION: INVOLVED IN OXYGEN TRANSPORT FROM THE LUNG TO THE
 CC VARIOUS PERIPHERAL TISSUES.
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA CHAINS AND TWO BETA CHAINS.
 CC -1- TISSUE SPECIFICITY: RED BLOOD CELLS.
 CC -1- MASS SPECTROMETRY: MW=15978; MW_ERR=0.2; METHOD=Electrospray.
 CC -1- MISCELLANEOUS: THIS FISH HAS TWO HEMOGLOBINS: CATHODIC AND ANODIC.
 CC THE CATHODIC HB AND ANODIC HB DISPLAY SMALL AND LARGE BOHR EFFECTS
 CC RESPECTIVELY. IN ADDITION, THE CATHODIC HB DISPLAYS A REVERSE BOHR
 CC EFFECT AND APPRECIABLE PHOSPHATE EFFECTS.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
 DR HSSP: P02142; 10UF
 DR InterPro: IPR002337; Beta_haem.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin.1.
 DR PRINTS: PR00814; BETAHAEM.
 DR PROSITE: PS01033; GLOBIN.1.
 KW Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 63 63
 FT METAL 92 92
 FT METAL 92 92
 SQ SEQUENCE 146 AA; 15976 MW; 4D75EB9FC8D73539 CRC64;
 QY 35 DATAI 40
 DB 9 DATAI 14
 Query Match 2.8%; Score 6; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 1.le+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 47
 Y142.METUA STANDARD; PRT; 147 AA.
 AC 057607;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0142.
 GN MJ0142.
 GN Methanococcus jannaschli.
 OS Methanococcus jannaschli.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weissink K.G., Merrick J.M., Glodok A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Kleen H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschli*.";
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: STRONG, TO M. JANNASCHII MJ0605.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

QY      4 FLIIIA 9
        |||||
Db       6 FLIIIA 11

RESULT 43
Y13K_BP74
AC      Y13K_BP74      STANDARD;      PRT;      117 AA.
P39504;
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Hypothetical 13.1 kDa protein in pser-alc intergenic region.
GN      Y13K OR PSET 3.
OS      Bacteriophage T4.
OC      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC      T4-like viruses.
OC      NCBI_TaxId=10665;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Kuter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA      Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT      "Bacteriophage T4 genome analysis.";
RL      Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: AF158101; AAD42645.1; -
KW      Hypothetical protein.
SQ      SEQUENCE 117 AA; 13135 MW; D8B52577D7DA979B CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      36 AIAAIE 41
        |||||
Db       69 AIAAIE 74

RESULT 44
GCSH_THEMA
ID      GCSH_THEMA      STANDARD;      PRT;      124 AA.
AC      Q9WY55;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Probable glycine cleavage system H protein.
GN      GCVH OR TM0212.
OS      Thermotoga maritima.
OC      Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC      Thermotogaceae; Thermotoga.
OC      NCBI_TaxId=2336;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=99287316; PubMed=10360571;
RA      Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA      Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA      McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA      Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA      Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA      Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT      "Evidence for lateral gene transfer between Archaea and Bacteria from
RT      Nature 399:323-329(1999).";
RL      Nature 399:323-329(1999).
-1- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYZES THE DEGRADATION OF

```

```

CC      GLYCINE. THE H PROTEIN SHUTTLES THE METHYLAMINE GROUP OF GLYCINE
CC      FROM THE P PROTEIN TO THE T PROTEIN (BY SIMILARITY).
CC      -1- COFACTOR: THE H CHAIN CONTAINS A COVALENTLY-BOUND LIPOYL COFACTOR
CC      (BY SIMILARITY).
CC      -1- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:
CC      P, T, L, AND H (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE GCVH FAMILY.
CC      -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: AE001706; AAD35304.1; -
DR      HSSP: P16048; 1H7P.
DR      TIGR: TM0212; -
DR      InterPro: IPR002930; GCV_H.
DR      InterPro: IPR003016; LIPOYL.
DR      Pfam: PF01597; GCV_H; 1.
DR      TIGRFAMs: TIGR00527; gcvH; 1.
DR      PROSITE: PS00189; LIPOYL; 1.
KW      LIPOYL; Complete proteome.
FT      BINDING 60
SQ      SEQUENCE 124 AA; 13915 MW; 90B7CDE430A16C9F CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      145 EISDEG 150
        |||||
Db       101 EISDEG 106

RESULT 45
JANA_DROME
ID      JANA_DROME      STANDARD;      PRT;      135 AA.
AC      P20348; Q9VAB6;
DT      01-FEB-1991 (Rel. 17, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Sex-regulated protein Janus-A.
GN      JANA OR CG7933.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC      Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC      NCBI_TaxId=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Canton-S;
RX      MEDLINE=89343970; PubMed=2503707;
RA      Yancostas C., Vincent A., Lepesant J.-A.;
RT      "Transcriptional and posttranscriptional regulation contributes to
RT      the sex-regulated expression of two sequence-related genes at the
RT      Janus locus of Drosophila melanogaster.";
RL      Mol. Cell. Biol. 9:2526-2535(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Berkely;
RA      Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```



```

DR InterPro: IPR000032; HPr_protein.
DR Pfam: PF00381; PTS_HPr; 1.
DR PRINTS: PR00107; PHOSPHOCPPR.
DR TIGRFAms: TIGR01003; PTS_HPr_family; 1.
DR PROSITE: PS00589; PTS_HPr_HIS; 1.
DR PROSITE: PS00589; PTS_HPr_SER; 1.
KW Phosphotransferase system; Sugar transport; Phosphorylation.
FT MOD_RES 15 15 PHOSPHORYLATION (BY ENZYME 1) (BY
FT MOD_RES 46 46 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 88 AA; 9253 MW; 5723EBB1345F56CA CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 DATAI 40
Db 72 DATAI 77

RESULT 41
ES61_MYCTU STANDARD; PRT; 94 AA.
AC P96364; 008120; 008122; Q9L781;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Antigen MCB9.9B (ESAT-6 like protein 1).
GN (RV1037C OR MT1066 OR MTCY1062.12) AND
GN (RV3619C OR MT3721 OR MTCY15C10.33 OR MTCY07H7B.03).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Erkman;
RX MEDLINE=20130298; PubMed=10662800;
RA Alderson M.R., Bement T., Day C.H., Zhu L., Molesh D., Skelky Y.A.W.,
RA Coler R., Lewinson D.M., Reed S.G., Dillon D.C.,
RA "Expression cloning of an immunodominant family of Mycobacterium
RA tuberculosis antigens using human CD4(+) T cells.";
RL J. Exp. Med. 191:551-560(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Omayan L.A., Ertolovaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterlbeck T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF226277; AAF32406.1; -.
DR EMBL: Z92539; CAB06842.1; -.
DR EMBL: Z95436; CAB08822.1; -.
DR EMBL: AE006989; AAK45317.1; -.
DR EMBL: AE007171; AAK48082.1; -.
DR TIGR: MT1066; -.
DR TIGR: MT3721; -.
DR Tuberculist; RV1037C; -.
DR Tuberculist; RV3619C; -.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 20 20 Q -> L (IN REF. 1 AND 3; AAK48082).
FT CONFLICT 23 23 S -> L (IN REF. 1 AND 3; AAK48082).
SQ SEQUENCE 94 AA; 9833 MW; 8F971BB03789C57E CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SDVUTA 188
Db 33 SDVUTA 38

RESULT 42
NUIC_PLEBO STANDARD; PRT; 101 AA.
AC Q00244;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-plastoquinone oxidoreductase chain 4L (EC 1.6.5.3).
GN NDH4L.
OS Plectonema boryanum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.
OX NCBI_TaxId=1184;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM M-101;
RA Takahashi Y., Shonai F., Fujita Y., Kohchi T., Ohyama K.,
RA Matsudaira H.;
RA "Structure of a co-transcribed gene cluster, ndh1-frx8-ndh6-ndh4L,
RT cloned from the filamentous cyanobacterium Plectonema boryanum.";
RL Plant Cell Physiol. 32:969-981(1991).
CC -!- CATALYTIC ACTIVITY: NADH + plastoquinone = NAD(+) + plastoquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D01014; BAA00817.1; -.
DR PIR: JQ2138; JQ2138.
DR InterPro: IPR003215; NADH_dh_ubiqL.
DR InterPro: IPR001133; Oxidored_4L.
DR Pfam: PF00420; Oxidored_g2; 1.
DR ProDom: PD002107; NADH_dh_ubiq1; 1.
DR Oxidoreductase; Nad; Plastoquinone.
KW SEQUENCE 101 AA; 11335 MW; 9FD1A3B830A942F7 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92316212; PubMed-1618309;
RA Scaramuzzi C.D., Stokes H.W., Hittler R.G.;
RT "Characterisation of a chloroplast-encoded secY homologue and atpH
RT from a chromophytic alga. Evidence for a novel chloroplast genome
RT organisation."
RL FEBS Lett. 304:119-123(1992).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out)
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -1- MISCELLANEOUS: DICYCLOHEXYLCARBODIIMIDE (DCCD) INHIBITS ATPASE.
CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X64731; CAA45997.1; -.
CC PIR: S23424; S23424.
CC DR InterPro: IPR002379; ATPase_Csub.
CC DR InterPro: IPR000454; Eub_ATPase_Csub.
CC DR Pfam: PF00137; ATP-synt_C; 1.
CC DR PRINTS: PR00124; ATPASEC.
CC DR TIGRFS: TIGR01260; ATP_synt_C; 1.
CC DR PROSITE: PS00605; ATPASE_C; 1.
CC KW Hydrogen ion transport; Lipid-binding; Chloroplast; CF(0);
CC Transmembrane.
CC FT BINDING 61 61 DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).
CC SQ SEQUENCE 83 AA; 8180 MW; B41F6C0ABA086493 CRC64;

Query Match
Best Local Similarity 2.8%; Score 6; DB 1; Length 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VVALSL 144
Db 69 VVALSL 74

RESULT 39
YK45_PSEAE STANDARD; PRT; 86 AA.
AC 091270;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein PA2045.
GN PA2045.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Iarls K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen."

```

```

RL Nature 406:959-964(2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0161 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE004631; AAG05433.1; -.
CC DR InterPro: IPR002696; DUF37.
CC DR Pfam: PF01809; DUF37; 1.
CC DR ProDom: PD004225; DUF37; 1.
CC DR TIGRFS: TIGR00278; DUF37; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 86 AA; 9824 MW; 5E228B8ED7F5A829 CRC64;

Query Match
Best Local Similarity 2.8%; Score 6; DB 1; Length 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MKFLLI 7
Db 1 MKFLLI 6

RESULT 40
PTPH_LACCA STANDARD; PRT; 88 AA.
AC 09KJV3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphocarrier protein Hpr (Histidine-containing protein).
GN PTH.
OS Lactobacillus casei.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1582;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 393;
RX MEDLINE-20305547; PubMed-10844647;
RA Viana R., Monedero V., Dossonnet V., Vadeboncoeur C.,
RA Perez-Martinez G., Deutscher J.;
RT "Enzyme I and Hpr from Lactobacillus casei: their role in sugar
RT transport, carbon catabolite repression and inducer exclusion."
RL Mol. Microbiol. 36:570-584(2000).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE
CC TRANSPORT SYSTEM. THE PHOSPHORYL GROUP FROM PHOSPHOENOLPYRUVATE
CC (PEP) IS TRANSFERRED TO THE PHOSPHORYL CARRIER PROTEIN HPR BY
CC ENZYME I. PHOSPHO-HPR THEN TRANSFERS IT TO THE PERMEASE (ENZYMES
CC II/III). HPR IS COMMON TO ALL PTS (BY SIMILARITY).
CC -1- ENZYME REGULATION: PHOSPHORYLATION ON SER-46 INHIBITS THE
CC PHOSPHORYL TRANSFER FROM ENZYME I TO HPR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE HPR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF159589; AAF74346.1; -.
CC DR HSSP: P07515; IPTF.
CC DR InterPro: IPR001020; HPr_HisP_site.
CC DR InterPro: IPR002114; HPr_SerP_site.

```

Db 69 VVALSL 74

RESULT 36
ATPH_PORPU STANDARD: PRT: 82 AA.

AC P51246;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase C chain (EC 3.6.3.14) (Lipid-binding protein) (Subunit III).
GN ATPH.
OS Porphyra purpurea.
OC Chloroplast.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Reith M.E., Munholland J.;
RT *Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.;
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate + H(+) (out).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0) HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -1- MISCELLANEOUS: DICYCLOHEXYLCARBODIIMIDE (DCCD) INHIBITS ATPASE.
CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U38804; AAC08132.1; -
DR InterPro: IPR002379; ATPase_Csub.
DR InterPro: IPR000454; Eub_ATPase_Csub.
DR Pfam: PF00137; ATP-synt_C; 1.
DR PRINTS: PR00124; ATPSEC.
DR TIGRFS: TIGR01260; ATP-synt_c; 1.
DR PROSITE: PS00605; ATPASE_C; 1.
KM Hydrogen ion transport; Lipid-binding; Chloroplast; CF(0);
KW Transmembrane.
FT TRANSMEM 7 33
FT BINDING 51 77
FT SEQUENCE 82 AA; 8084 MW; 295BD8FA52C62D81 CRC64;
SQ

Query Match 2.8%; Score 6; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 VVALSL 144
Db 69 VVALSL 74

RESULT 37
ATPH_GALSU STANDARD: PRT: 83 AA.
AC P35013;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ATP synthase C chain (EC 3.6.3.14) (Lipid-binding protein) (Subunit III).
GN ATPH.
OS Galdieria sulphuraria.
OC Chloroplast.
OC Eukaryota; Bangiophyceae; Porphyridiales; Porphyridiaceae; Galdieria.
OX NCBI_TaxID=130081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=14-1 / Isolate 107.79/Goettingen;
RX MEDLINE=94033298; PubMed=8219057;
RA Kostzewa M., Zetsche K.;
RT *Organization of plastid-encoded ATPase genes and flanking regions including homologues of infb and tsf in the thermophilic red alga Galdieria sulphuraria.;
RL Plant Mol. Biol. 23:67-76(1993).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate + H(+) (out).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0) HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -1- MISCELLANEOUS: DICYCLOHEXYLCARBODIIMIDE (DCCD) INHIBITS ATPASE.
CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X67814; CAA48021.1; -
DR PIR: S39516; S39516.
DR InterPro: IPR002379; ATPase_Csub.
DR InterPro: IPR000454; Eub_ATPase_Csub.
DR Pfam: PF00137; ATP-synt_C; 1.
DR PRINTS: PR00124; ATPSEC.
DR TIGRFS: TIGR01260; ATP-synt_c; 1.
DR PROSITE: PS00605; ATPASE_C; 1.
KM Hydrogen ion transport; Lipid-binding; Chloroplast; CF(0);
KW Transmembrane.
FT TRANSMEM 7 33
FT BINDING 51 77
FT SEQUENCE 83 AA; 8282 MW; CEC80894FB998E76 CRC64;
SQ

Query Match 2.8%; Score 6; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 VVALSL 144
Db 69 VVALSL 74

RESULT 38
ATPH_PAVLU STANDARD: PRT: 83 AA.
AC P28530;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase C chain (EC 3.6.3.14) (Lipid-binding protein) (Subunit III).
GN ATPH.
OS Pavlova lutheri (Monochrysis lutheri).
OC Chloroplast.
OC Eukaryota; Haptophyceae; Pavlovaes; Pavlova.
OX NCBI_TaxID=2832;

DR InterPro: IPR003231; Acyl-carrier.
 DR InterPro: IPR003880; Pantine-attach.
 DR Pfam: PF00550; pp-binding; 1.
 DR ProDom: PD000887; Acyl-carrier; 1.
 DR TIGRfam: TIGR00517; acyl-carrier; 1.
 DR PROSITE: PS50075; ACP_DOMAIN; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
 DR Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;
 KW Complete proteome.
 FT BINDING 37 37 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 SQ SEQUENCE 77 AA; 8558 MM; 3575CDFA5BA15CA CRC64;
 Query Match 2.8%; Score 6; DB 1; Length 77;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 144 LEISDE 149
 DB 53 LEISDE 58
 RESULT 34
 ATPH_ANTSP STANDARD; PRT; 82 AA.
 ID ATPH_ANTSP
 AC Q02851;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP synthase C chain (EC 3.6.3.14) (Lipid-binding protein) (Subunit
 DE I11).
 GN ATPH.
 OS Anthamion sp.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiales;
 OC Anthamion.
 OX NCBI_TaxID=2767;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LB 95.79;
 RX MEDLINE-93021132; PubMed-1404401;
 RA Koziczewa M., Zetsche K.;
 RT "Large ATP synthase operon of the red alga Anthamion sp. resembles
 RL the corresponding operon in cyanobacteria.";
 RL J. Mol. Biol. 227:961-970(1992).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
 CC H(+)(Out).
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -1- MISCELLANEOUS: DICYCLOHEXYLCARBODIIMIDE (DCCD) INHIBITS ATPASE.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: XG3382; CAA44980.1; -.
 DR PIR: S26958; S26958.
 DR InterPro: IPR002379; ATPase_Csub.
 DR InterPro: IPR000454; Eub_ATPase_Csub.
 DR Pfam: PF00137; ATP-synl_C; 1.
 DR PRINTS: PR00124; ATPASEC.
 DR TIGRfam: TIGR01260; ATP-synl_C; 1.
 DR PROSITE: PS00605; ATPASE_C; 1.
 DR Hydrogen ion transport; Lipid-binding; Chloroplast; CF(0);
 KW Transmembrane.
 FT TRANSMEM 7 33 BY SIMILARITY.

FT TRANSMEM 51 77 BY SIMILARITY.
 FT BINDING 61 61 DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).
 SQ SEQUENCE 82 AA; 8100 MM; 3F22042A52DE9622 CRC64;
 Query Match 2.8%; Score 6; DB 1; Length 82;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 139 VVALSL 144
 DB 69 VVALSL 74
 RESULT 35
 ATPH_CYACA STANDARD; PRT; 82 AA.
 ID ATPH_CYACA
 AC Q9TM30;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP synthase C chain (EC 3.6.3.14) (Lipid-binding protein) (Subunit
 DE I11).
 GN ATPH.
 OS Cyanidium caldarium.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
 OC Cyanidium.
 OX NCBI_TaxID=2771;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RK-1;
 RX MEDLINE-20496959; PubMed-11040290;
 RA Gloeckner G., Rosenthal A., Valentin K.-U.;
 RT "The structure and gene repertoire of an ancient red algal plastid
 RT genome.";
 RL J. Mol. Evol. 51:382-390(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
 CC H(+)(Out).
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -1- MISCELLANEOUS: DICYCLOHEXYLCARBODIIMIDE (DCCD) INHIBITS ATPASE.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF022186; AAF13009.1; -.
 DR InterPro: IPR002379; ATPase_Csub.
 DR InterPro: IPR000454; Eub_ATPase_Csub.
 DR Pfam: PF00137; ATP-synl_C; 1.
 DR PRINTS: PR00124; ATPASEC.
 DR TIGRfam: TIGR01260; ATP-synl_C; 1.
 DR PROSITE: PS00605; ATPASE_C; 1.
 DR Hydrogen ion transport; Lipid-binding; Chloroplast; CF(0);
 KW Transmembrane.
 FT TRANSMEM 7 33 POTENTIAL.
 FT TRANSMEM 51 77 POTENTIAL.
 FT BINDING 61 61 DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).
 SQ SEQUENCE 82 AA; 8149 MM; 908E74D1BB0137CF CRC64;
 Query Match 2.8%; Score 6; DB 1; Length 82;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 139 VVALSL 144

QY 10 VAEVAV 15
 |||||
 DB 26 VAEVAV 31

RESULT 32
 A95E.DROME
 ID A95E.DROME STANDARD: PRT: 52 AA.
 AC p16548: O9VNB0:
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Accessory gland-specific peptide 95EF precursor (Male accessory gland secretory protein 316).
 GN ACP95EF OR MSP95E OR MSP316 OR CGI7924.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G., Abril J.F., Aghayani A., An H.-J., Andrews-Planckoch C., Baldwin D., Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C., Jatalin M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moperson D., Merklow G., Mishina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacled J.M., Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spter E., Spreading A.C., Stapleton M., Strong R., Sun E., Syntekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:215-2195(2000).
 CC -1- FUNCTION: THIS PROTEIN MAY BE A PRECURSOR OF SECRETED PROTEINS
 CC AND PEPTIDE HORMONES.

CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: MAIN CELLS OF THE ACCESSORY GLANDS OF MALES.
 CC -1- DEVELOPMENTAL STAGE: IN VERY LATE PUPAE AND IN ADULTS.
 CC -1- INDUCTION: BY MATING.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M32022; AAD13400.1; -
 CC DR EMBL: AEO03746; AAF56263.2; -
 CC DR FlyBase: Fggn0002863; ACP95EF.
 CC KW Behavior: Hormone; Signal.
 CC FT SIGNAL 1 22 POTENTIAL.
 CC FT CHAIN 23 52 ACCESSORY GLAND-SPECIFIC PEPTIDE 95EF.
 CC FT CONFLICT 20 20 T -> S (IN REF. 1).
 CC SQ SEQUENCE 52 AA; 5405 MW; AA4008BD6057FE2B CRC64;

Query Match 2.8%; Score 6; DB 1; Length 52;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VVALSL 144
 |||||
 DB 13 VVALSL 18

RESULT 33
 ACP_BACHD
 ID ACP_BACHD STANDARD: PRT: 77 AA.
 AC 09KA04;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Acyl carrier protein (ACP).
 GN ACP OR ACPA OR BH2490.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512562; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N., Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S., Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*.";
 RT Nucleic Acids Res. 28:4317-4331(2000).
 RL
 CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty acid biosynthesis (By similarity).
 CC -1- PATHWAY: De novo fatty acid biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- PTM: 4'-phosphopantetheine is transferred from CoA to a specific serine of apo-ACP by acps. This modification is essential for activity because fatty acids are bound in thioester linkage to the sulphydryl of the prosthetic group (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AP001515; BAB06209.1; -
 CC DR HSSP: P80643; 1HY8.

CC Bacteria; Proteobacteria; gamma subdivision; Ectothiorhodospiraceae;
 OC Halorhodospira.
 OX NCBI_TaxID=1052;
 RN [1]
 RP SEQUENCE.
 RC STRAIN-DSM 1059;
 RX MEDLINE-92249336; PubMed-1577009;
 RA Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
 Zuber H.;
 RT "The primary structure of the antenna polypeptides of
 Ectothiorhodospira halochloris and Ectothiorhodospira halophila. Four
 core-type antenna polypeptides in E. halochloris and E. halophila."
 RL Eur. J. Biochem. 205:917-925(1992).
 CC -1- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
 TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
 CC -1- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
 CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
 CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
 CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
 CC ADDITIONAL COMPONENTS.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
 CC PIR: S23286; S23286.
 DR HSSP: P02951; 1DX7.
 DR InterPro: IPR002362; Antenna_beta.
 DR PROSITE: PS00969; ANTENNA_COMP_BETA; PARTIAL.
 KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
 KW Magnesium; Bacteriochlorophyll; Inner membrane.
 FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 21 >40 POTENTIAL.
 FT BINDING 19 19 BACTERIOCHLOROPHYLL (POTENTIAL).
 FT METAL 37 37 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
 MAGNESIUM (POTENTIAL).
 FT NON_TER 40 40
 SQ SEQUENCE 40 AA; 4575 MW; 63536682D22B5D7C CRC64;
 Query Match 2.8%; Score 6; DB 1; Length 40;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 AVAFVA 14
 DB 31 AVAFVA 36

RESULT 30
 LHB1_RHOTE STANDARD; PRT; 48 AA.
 ID LHB1_RHOTE
 AC P80590;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Light-harvesting polypeptide B-885, beta-1 chain (LH-1) (Antenna
 DE pigment polypeptide, beta-1 chain).
 OS Rhodocyclus tenuis (Rhodospirillum tenue).
 CC Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;
 CC Rhodocyclus.
 OX NCBI_TaxID=1066;
 RN [1]
 RP SEQUENCE.
 RC STRAIN-DSM 109;
 RX MEDLINE-96283832; PubMed-8681949;
 RA Hu Q., Brunisholz R.A., Frank G., Zuber H.;
 RT "The antenna complexes of the purple non-sulfur photosynthetic
 bacterium Rhodocyclus tenuis. Structural and spectral
 characterization."
 RL Eur. J. Biochem. 238:381-390(1996).
 CC -1- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
 TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
 CC -1- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
 CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
 CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
 CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
 CC ADDITIONAL COMPONENTS.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
 DR HSSP: P02951; 1DX7.
 DR InterPro: IPR000066; Antenna_a/b.
 DR InterPro: IPR002362; Antenna_beta.
 DR Pfam: PF00556; LHC; 1.
 DR PRINTS: PR00674; LIGHTHARVSTB.
 DR PROSITE: PS00969; ANTENNA_COMP_BETA; FALSE_NEG.
 KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
 KW Magnesium; Bacteriochlorophyll; Inner membrane.
 FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 21 43 PERIPLASMIC (POTENTIAL).
 FT DOMAIN 44 48 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
 FT METAL 37 37 MAGNESIUM (POTENTIAL).
 SQ SEQUENCE 48 AA; 5344 MW; 8CB2F1C3736F0086 CRC64;
 Query Match 2.8%; Score 6; DB 1; Length 48;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 VAFVAV 15
 DB 26 VAFVAV 31

RESULT 31
 LHB2_RHOTE STANDARD; PRT; 48 AA.
 ID LHB2_RHOTE
 AC P80591;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Light-harvesting polypeptide B-885, beta-2 chain (LH-1) (Antenna
 DE pigment polypeptide, beta-2 chain).
 OS Rhodocyclus tenuis (Rhodospirillum tenue).
 CC Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;
 CC Rhodocyclus.
 OX NCBI_TaxID=1066;
 RN [1]
 RP SEQUENCE.
 RC STRAIN-DSM 109;
 RX MEDLINE-96283832; PubMed-8681949;
 RA Hu Q., Brunisholz R.A., Frank G., Zuber H.;
 RT "The antenna complexes of the purple non-sulfur photosynthetic
 bacterium Rhodocyclus tenuis. Structural and spectral
 characterization."
 RL Eur. J. Biochem. 238:381-390(1996).
 CC -1- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
 TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
 CC -1- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
 CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
 CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
 CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
 CC ADDITIONAL COMPONENTS.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
 CC HSSP: P02951; 1DX7.
 DR InterPro: IPR000066; Antenna_a/b.
 DR InterPro: IPR002362; Antenna_beta.
 DR Pfam: PF00556; LHC; 1.
 DR PRINTS: PR00674; LIGHTHARVSTB.
 DR PROSITE: PS00969; ANTENNA_COMP_BETA; FALSE_NEG.
 KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
 KW Magnesium; Bacteriochlorophyll; Inner membrane.
 FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 21 43 PERIPLASMIC (POTENTIAL).
 FT DOMAIN 44 48 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
 FT METAL 37 37 MAGNESIUM (POTENTIAL).
 SQ SEQUENCE 48 AA; 5358 MW; 8CB2F1C3736F1825 CRC64;
 Query Match 2.8%; Score 6; DB 1; Length 48;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

SYA_BARBA
ID SYA_BARBA STANDARD: PRT; 886 AA.
AC P70865;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alanyl-L-cRNA synthetase (EC 6.1.1.7) (Alanine--cRNA ligase) (Alars).
GN ALAS.
OS Bartonella bacilliformis.
OC Bacteriia; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxId=774;
RN [1]
RP SEQUENCE FROM N.A.
RA Upeslacs E., Ihler G.M.;
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC di-phosphate + L-alanyl-tRNA(Ala).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; U68242; AAB09037.1; -
DR InterPro: IPR002106; ATRNA_LigaseII.
DR InterPro: IPR003156; DHNA1.
DR InterPro: IPR002318; tRNA-synt_2c.
DR Pfam: PF01411; tRNA-synt_2c; 1.
DR Pfam: PF02272; DHNA1; 1.
DR PRINTS: PRO0980; TRNASYNTHALA.
DR TIGRFA: TIGR00344; alas; 1.
DR PROSITE: PS50860; AA_tRNA_Ligase_II_ALA; 1.
KW Aminoacyl-L-cRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ SEQUENCE 886 AA; 97907 MW; 845AADB4611BCA42 CRC64;

Query Match 3.3%; Score 7; DB 1; Length 886;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 AIDDATA 38
Db 632 AIDDATA 638

RESULT 28
SECT_YEAST STANDARD: PRT; 2009 AA.
ID SECT_YEAST STANDARD: PRT; 2009 AA.
AC P11075; O03960; O04139;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein transport protein SECT7.
GN SECT7 OR YDR170C OR YD9395.01C OR YD9489.05C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Achsletter T., Franzusoff A., Field C., Schekman R.;
RL "SECT7 encodes an unusual, high molecular weight protein required for
RL membrane traffic from the yeast Golgi apparatus.";
RL J. Biol. Chem. 263:11711-11717(1988).
RN [2]
RP SEQUENCE OF 1-542 FROM N.A.
RC STRAIN=S288c / AB972;

```

```

RA Murphy L., Harris D.E., Barrell B.G., Rajandream M.A.;
RL Submitted (Nov-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 204-2009 FROM N.A.
RC STRAIN=S288c / AB972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;
RL Submitted (Jan-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91093340; PubMed=1986005;
RA Franzusoff A., Redding K., Crosby J., Fuller R.S., Schekman R.;
RT "Localization of components involved in protein transport and
RT processing through the yeast Golgi apparatus.";
RL J. Cell Biol. 112:27-37(1991).
CC -1- FUNCTION: MAY PLAY A ROLE IN VESICULAR BUDDING AND TRAFFIC BETWEEN
CC COMPARTMENTS OF THE GOLGI APPARATUS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; ASSOCIATED WITH THE PERIPHERAL
CC GOLGI MEMBRANE.
CC -1- DOMAIN: THE HIGHLY CHARGED ACIDIC DOMAIN MAY SERVE A STRUCTURAL
CC ROLE TO INTERACT WITH LIPIDS OR PROTEINS ON THE CYTOPLASMIC
CC SURFACE OF THE GOLGI APPARATUS.
CC -1- PTM: PHOSPHORYLATED.
CC -1- SIMILARITY: CONTAINS 1 SECT7 DOMAIN.
CC -1- SIMILARITY: TO YEAST YEL022W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; J03918; AAB04031.1; -
DR EMBL; 246727; CAA86696.1; -
DR EMBL; 247813; CAA87801.1; -
DR PIR; A28784; A28784.
DR PIR; A31068; A31068.
DR HSSP; Q99418; 1PBV.
DR SGD; S0002577; SECT7.
DR InterPro: IPR000904; SECT7.
DR Pfam: PF01369; SECT7; 1.
DR SMART; SM00222; SECT7; 1.
DR PROSITE; PS50190; SECT7; 1.
KW Transport; Protein transport; Golgi stack; Phosphorylation.
FT DOMAIN 89 213 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 824 1010 SECT7.
FT CONFLICT 188 188 A -> S (IN REF. 1).
FT CONFLICT 399 400 FV -> L (IN REF. 1).
FT CONFLICT 402 402 S -> C (IN REF. 1).
FT CONFLICT 1031 1034 QOSA -> PAIC (IN REF. 1).
FT CONFLICT 1036 1037 NF -> QL (IN REF. 1).
SQ SEQUENCE 2009 AA; 226885 MW; 02BD2370DD2E4661 CRC64;

Query Match 3.3%; Score 7; DB 1; Length 2009;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 PIRGVLS 183
Db 1588 PIRGVLS 1594

RESULT 29
LHBL_ECTHL STANDARD: PRT; 40 AA.
ID LHBL_ECTHL STANDARD: PRT; 40 AA.
AC P80106;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Light harvesting protein B800/830/1020, beta-1 chain (EHS-beta-1)
DE (Antenna pigment protein, beta-1 chain) (Fragment).
OS Ectochlorodospira halochloris.

```

CC STRAIN=PR745;
 RX MEDLINE=98162722; PubMed=9501991;
 RA Yoshioaka S., Kato K., Nakai K., Okayama H., Nojima H.;
 RT "Identification of open reading frames in Schizosaccharomyces pombe
 RT CDMS-";
 RL DNA Res. 4:363-369(1997).
 CC -1- SIMILARITY: CONTAINS 5 WD_REPEATS (TRP-ASP DOMAINS).
 CC -1- SIMILARITY: BELONGS TO THE CGI-48 FAMILY OF WD-REPEAT PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AL022299; CAA1383.1; -;
 DR EMBL: D89098; BAA13761.1; -;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 4.
 DR SMART: SM00320; WD40; 4.
 DR PROSITE: PS00678; WD_REPEATS_1; FALSE_NEG.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Hypothetical protein; Repeat; WD repeat.
 FT REPEAT 216 254 WD 1.
 FT REPEAT 259 298 WD 2.
 FT REPEAT 390 429 WD 3.
 FT REPEAT 438 479 WD 4.
 FT REPEAT 485 523 WD 5.
 SQ SEQUENCE 556 AA; 62715 MW; 2DF02D921F9D5D7D CRC64;
 Query Match 3.3%; Score 7; DB 1; Length 556;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 GVLSDVL 186
 DB 350 GVLSDVL 356

RESULT 26
 VIRAL AGRT5 STANDARD; PRT; 833 AA.
 ID VIRAL AGRT5 STANDARD; PRT; 833 AA.
 AC P18540; G52297;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Wide host range VIRAL protein (EC 2.7.3.-) (VHR VIRAL).
 GN VIRAL OR ATU6166 OR AGR_PTL2.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Plasmid pTIC58.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90301800; PubMed=2194232;
 RA Rogowsky P.M., Powell B.S., Shirasu K., Lin T.-S., Morel P.,
 RA Zyrjan E.M., Steck T.R., Kado C.I.;
 RT "Molecular characterization of the vir regulation of Agrobacterium of
 RT tumefaciens: complete nucleotide sequence and gene organization of
 RT the 28.63-kbp region cloned as a single unit."
 RL Plasmid 23:85-106(1990).
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Powell G.K.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,

RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. St.,
 RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Senphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58."
 RL Science 294:2317-2323(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,
 RA Hounzel K., Goldman B.S., Cao Y., Askenazi M., Halling E., Liu F.,
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RT Science 294:2323-2328(2001).
 CC -1- FUNCTION: ACTIVATES VIRG, BY PHOSPHORYLATING IT, IN THE
 CC PRESENCE OF ACETOSYRINGONE OR HYDROXYSYRINGONE.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: J03320; AAA91590.1; -;
 DR EMBL: L48210; AAA9282.1; -;
 DR EMBL: AE009435; AAL46402.1; -;
 DR EMBL: AE007923; AAK90927.1; -;
 DR PIR: S11825; S11825.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR004358; Bact_sens_pr_C.
 DR InterPro: IPR004359; His_kin_sig.
 DR InterPro: IPR003661; His_kinA.
 DR InterPro: IPR001789; Response_reg.
 DR Pfam: PF00512; Signal; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR PRINTS: PR00344; BCTRLENSOR.
 DR PRODOM: PD000039; Response_reg; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00388; HIS_KIN; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 KW Sensory transduction; Transferrase; Kinase; Phosphorylation; Plasmid;
 KW Crown gall tumor; Transmembrane; Complete proteome.
 FT TRANSMEM 19 37
 FT DOMAIN 475 698 HISTIDINE KINASE.
 FT MOD_RES 478 478 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CONFLICT 68 68 R -> S (IN REF. 1).
 FT CONFLICT 322 322 L -> V (IN REF. 1).
 FT CONFLICT 455 455 I -> F (IN REF. 1).
 SQ SEQUENCE 833 AA; 91322 MW; B514DAR85DFB2B5 CRC64;
 Query Match 3.3%; Score 7; DB 1; Length 833;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 ALSLEIS 147
 DB 181 ALSLEIS 187

RESULT 27

Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 169 ICGSIL 175
|111111|
Db 90 ICGSIL 96

RESULT 24
CADL_DROME STANDARD; PRT; 511 AA.
AC Q9VLZ7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable cytochrome P450 4d21 (EC 1.14.-.-) (CYP4D21).
GN CYP4D21 OR CG6730.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burks K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin E., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei T., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacible J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zaveri J.S., Zhou M., Zhang Q., Zhao Q., Zheng X.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT The genome sequence of *Drosophila melanogaster*.
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE003618; AAF52531.1;
CC FLYBase: Fggn0031925; Cyp4d21.
CC InterPro: IPR001128; Cytochrome_P450.
CC Pfam: PF00067; P450_1.
CC PRINTS: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450; 1.
CC KMW Oxioreductase; Monooxygenase; Membrane; Heme; Microsome;
CC Endoplasmic reticulum; Hypothetical protein.
CC FT BINDING 456 456 HEME (BY SIMILARITY).
CC SEQUENCE 511 AA; 58264 MW; DF130DC603BDA50 CRC64;
CC -----
CC Query Match 3.3%; Score 7; DB 1; Length 511;
CC Best Local Similarity 100.0%; Pred. No. 31;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 KVLAPAF 206
|111111|
Db 129 KVLAPAF 135

RESULT 25
CG48_SCHPO STANDARD; PRT; 556 AA.
ID CG48_SCHPO
AC P78750;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein C29A3.06 in chromosome II.
GN SPBC29A3.06.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; Pubmed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felkewell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren V., Whitehead S.,
RA Woodward J., Volkhardt G., Aert R., Robben J., Grymoprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel K., Fuchs M., Fitz C., Holzer E., Mostl D., Hiltner H.,
RA Bazerly K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Beger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadiou E., Dreano S., Gloux S., Lelure V., Motiller S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochat M., Gallardin C., Talada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT The genome sequence of *Schizosaccharomyces pombe*.
RL Nature 415:871-880(2002).
CC [2]
CC SEQUENCE OF 269-556 FROM N.A.
RN RP

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KFLIIA 9
|||||||
Db 326 KFLIIA 332

RESULT 22
CHIO_HUMAN
ID CHIO_HUMAN STANDARD: PRT; 468 AA.
AC P52757;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-chimaerin (Beta-chimerin).
GN CHN2 OR BCH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RX MEDLINE=94230370; PubMed=8175705;
RA Leung T., How B.-E., Manser E., Lim L.;
RT "Cerebellar beta 2-chimaerin, a GTPase-activating protein for p21
RT ras-related rac is specifically expressed in granule cells and has a
RT unique N-terminal SH2 domain."
RL J. Biol. Chem. 269:12888-12892(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95339337; PubMed=7614486;
RA Yuan S., Miller D.W., Barnett G.H., Hahn J.F., Williams B.R.G.;
RT "Identification and characterization of human beta 2-chimerin;
RT association with malignant transformation in astrocytoma."
RL Cancer Res. 55:3456-3461(1995).
RN [3]
RP SEQUENCE OF 18-192 FROM N.A.
RA Strong C., Graves T., Yoakum M., Hawkins M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBD databases.
CC -1- FUNCTION: GTPASE ACTIVATING PROTEIN FOR P21-RAC. INSUFFICIENT
CC EXPRESSION OF BETA-2 CHIMAERIN IS EXPECTED TO LEAD TO HIGHER RAC
CC ACTIVITY AND COULD THEREFORE PLAY A ROLE IN THE PROGRESSION FROM
CC LOW-GRADE TO HIGH-GRADE TUMORS.
CC -1- SUBCELLULAR LOCATION: Membrane-associated (potential).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; BETA-1 AND BETA-2 (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN THE BRAIN AND PANCREAS. ALSO
CC EXPRESSED IN THE HEART, PLACENTA, AND WEAKLY IN THE KIDNEY AND
CC LIVER. EXPRESSION IS MUCH REDUCED IN THE MALIGNANT GLIOMAS,
CC COMPARED TO NORMAL BRAIN OR LOW-GRADE ASTROCYTOMAS.
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L29126; AAA19191.1; -
DR EMBL: U07223; AAA16836.1; -
DR EMBL: U28926; AAA6528.1; -
DR EMBL: AC004417; AAC06177.1; -
DR HSSP: P28867; 1PNO
DR Genew: HGNC:1944; CHN2.
DR MIM: 602857; -
DR InterPro: IPR002219; DAG_PE-bind.

DR InterPro: IPR000198; RhoGAP.
DR InterPro: IPR000980; SH2.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00130; DAG_PE-bind; 1.
DR Pfam: PF0620; RhoGAP; 1.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR ProDom: PD000093; SH2; 1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00324; RhoGAP; 1.
DR SMART: SM00252; SH2; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50001; SH2; 1.
KW GTPase activation; Phorbol-ester binding; zinc; SH2 domain;
KW Alternative splicing.
FT DOMAIN 59 127 SH2.
FT DOMAIN 215 264 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 291 435 RHO-GAP.
FT CONFLICT 1 6 MASSN -> MRL (IN REF. 1).
SQ SEQUENCE 468 AA; 53923 MW; 63254958E0B5804C CRC64;

Query Match 3.3%; Score 7; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 IEARGLK 82
|||||||
Db 301 IEARGLK 307

RESULT 23
ARAA_YERPE
ID ARAA_YERPE STANDARD: PRT; 500 AA.
AC P58540;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE L-arabinose isomerase (EC 5.3.1.4).
GN ARAA OR YPO2253.
OS Yersinia pestis.
OC Yersinia.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tittball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagsels K., Kariyasev A.V.,
RA Leather S., Moutle S., Oyston P.C.F., Quail M., Rutherford B.G.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
CC -1- CATALYTIC ACTIVITY: L-arabinose = L-ribulose.
CC -1- PATHWAY: L-arabinose catabolism: first step.
CC -1- SIMILARITY: BELONGS TO THE ARABINOSE ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ14151; CAC91057.1; -
KW Isomerase; Arabinose catabolism; Complete proteome.
SQ SEQUENCE 500 AA; 55598 MW; 44A2680B38FAA45 CRC64;

Query Match 3.3%; Score 7; DB 1; Length 500;

RP REVISIONS.
 RA Waterston R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOT KNOWN. PUTATIVE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U28740: AAA68317.3: -;
 DR Wormpep: B0563.6: CE28551.
 DR InterPro: IPR000276: GPCR_Rhodopsn.
 DR Pfam: PF00001: 7tm_1; 1.
 DR PROSITE: PS00237: G-PROTEIN_RECP_F1_1; 1.
 DR PROSITE: PS00262: G-PROTEIN_RECP_F1_2; 1.
 KW Hypothetical protein; G-protein coupled receptor; Transmembrane;
 KW Glycoprotein.
 FT TRANSMEM 30 50
 FT TRANSMEM 65 85
 FT TRANSMEM 105 125
 FT TRANSMEM 147 167
 FT TRANSMEM 208 228
 FT TRANSMEM 258 278
 FT CARBOHYD 12 12
 FT CARBOHYD 88 88
 FT CARBOHYD 181 181
 FT CARBOHYD 429 429
 FT CARBOHYD 430 430
 SQ SEQUENCE 434 AA: 50026 MW: D5936FD2470C01D CRC64;
 Query Match 3.3%; Score 7; DB 1; Length 434;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 139 VALSLE 145
 DB 121 VALSLE 127
 RESULT 21
 ORC5_DROME
 ID ORC5_DROME STANDARD; PRT; 460 AA.
 AC Q24169; Q9V398;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Origin recognition complex subunit 5.
 GN ORC5 OR CG7833.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-96099403; PubMed-7502079;
 RA Gossens M., Pak D.T.S., Hansen S.K., Acharya J.K., Botchan M.R.;
 RT Science 270:1674-1677(1995).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE-99403001; PubMed-10471707;
 RA Ashburner M., Mistra S., Roote J., Lewis S.E., Blazey R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hattzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitehead K.,

RA Celniker S., Rubin G.M.;
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region.";
 RL Genetics 153:179-219(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baydalle J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandal D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fester C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
 CC BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
 CC REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
 CC THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN
 CC ATP-DEPENDENT MANNER.
 CC -1- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ORC5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U43505; AAC46956.1: -;
 DR EMBL: AE003408; AAF44824.1: -;
 DR EMBL: AE003641; AAF53340.1: -;
 DR FlyBase: FBgn0015271; Orc5.
 KW DNA replication; Nuclear protein; ATP-binding.
 FT NP_BIND 41 48
 FT SEQUENCE 460 AA: 52115 MW: FDCE3969E1CB7D2 CRC64;
 Query Match 3.3%; Score 7; DB 1; Length 460;
 Best Local Similarity 100.0%; Pred. No. 28;

```

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Seemphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.,
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58." ;
RT Science 294:2317-2323(2001).
[4]
RN SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quicilo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houniel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
RA Mollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58." ;
RT Science 294:2323-2328(2001).
CC -1- CATALYTIC ACTIVITY: L-ornithine = L-proline + NH(3).
CC -1- COFACTOR: NAD(+).
CC -1- ENZYME REGULATION: ACTIVITY IS SUBJECT TO SUBSTRATE INHIBITION
CC AND IS REGULATED BY L-ARGININE.
CC -1- PATHWAY: Conversion of nopaline to proline; last step.
CC -1- SIMILARITY: TO OTHER ORNITHINE CYCLODEAMINASES.
CC -1- SIMILARITY: REGIONS OF SIMILARITY WITH E.COLI AND P.AERUGINOSA
CC CARBAMOYLTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; X07435; CAA30316.1; -
DR EMBL; Z30316; CAA82966.1; -
DR EMBL; AE009420; AAL46252.1; ALF_INIT.
DR EMBL; AE007928; AAK90974.1; ALF_INIT.
DR PIR; S00402; DUACO.
DR InterPro; IPR003462; ODC_Mu_crystall.
DR Pfam; PF02423; ODC_Mu_crystall; 1.
KW Lyase; NAD; Plasmid; Complete proteome.
FT CONFLICT 212 212 K -> E (IN REF. 1 AND 2).
FT CONFLICT 297 297 L -> I (IN REF. 1 AND 2).
SQ SEQUENCE 354 AA; 38983 MW; 8272024E0BFDAR24 CRC64;
Query Match 3.3%; Score 7; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 179 FGVLSDV 185
DB 96 FGVLSDV 102
RESULT 19
OCD_AGR4
ID OCD_AGR4 STANDARD; PRT; 356 AA.
AC Q59701;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ornithine cyclodeaminase (EC 4.3.1.12) (OCD).
GN ARCB OR OCD.
OS Agrobacterium tumefaciens (strain Ach5).

```

```

OG Plasmid pTiAch5.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCHI_TaxID=176298;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89123162; PubMed=2644238;
RA Schrodner U., Sans N., Schrodner J.;
RT "Ornithine cyclodeaminase from octopine Ti plasmid Ach5:
RT identification, DNA sequence, enzyme properties, and comparison with
RT gene and enzyme from nopaline Ti plasmid C58." ;
RT J. Bacteriol. 171:847-854(1989).
[2]
RN SEQUENCE FROM N.A.
RA Schrodner J., von Lintig J., Zanker H.;
RT "Catabolism of the guanidino compounds nopaline, octopine, and
RT L-arginine in Agrobacterium tumefaciens: enzymes, genes, and
RT regulation." ;
RL (in) De Deyn P.P., Marescau B., Stalon V., Qureshi I.A. (eds.);
RL Guanidino compounds in biology and medicine, pp.19-28, J. Libbey,
RL London (1992)
CC -1- CATALYTIC ACTIVITY: L-ornithine = L-proline + NH(3).
CC -1- COFACTOR: NAD(+).
CC -1- ENZYME REGULATION: ACTIVITY IS SUBJECT TO SUBSTRATE INHIBITION
CC AND IS REGULATED BY L-ARGININE.
CC -1- PATHWAY: Conversion of nopaline to proline; last step.
CC -1- SIMILARITY: TO OTHER ORNITHINE CYCLODEAMINASES.
CC -1- SIMILARITY: REGIONS OF SIMILARITY WITH E.COLI AND P.AERUGINOSA
CC CARBAMOYLTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; M24146; AAA50518.1; -
DR EMBL; Z30328; CAA82989.1; -
DR InterPro; IPR003462; ODC_Mu_crystall.
DR Pfam; PF02423; ODC_Mu_crystall; 1.
KW Lyase; NAD; Plasmid.
SQ SEQUENCE 356 AA; 39166 MW; 5B3D3765E4800395 CRC64;
Query Match 3.3%; Score 7; DB 1; Length 356;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 179 FGVLSDV 185
DB 96 FGVLSDV 102
RESULT 20
YTF6_CAEEL
ID YTF6_CAEEL STANDARD; PRT; 434 AA.
AC Q11082;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor B0563.6.
GN B0563.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBL_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RA Faveillo T.;
RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[2]

```

```

CC -----
DR EMBL: X69489; CAA49244.1; -
DR EMBL: L07494; AAA40809.1; -
DR HSSP: P28867; IPTO.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR000198; RhogAP.
DR Pfam: PF00130; DAG_PE-bind; 1.
DR Pfam: PF00620; RhogAP; 1.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR SMART: SM00109; C1; 1.
DR SMART: SM00324; RhogAP; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS00081; DAG_PE_BIND_DOM_2; 1.
DR GTPase activation; Phorbol-ester binding; zinc; Membrane; SH2 domain;
KW Alternative splicing.
FT DOMAIN 42 91 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 118 262 RHO-GAP.
SQ SEQUENCE 295 AA; 33837 MW; D7692D957B4816BD CRC64;

Query Match 3.3%; Score 7; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 IEARGIK 82
DB 128 IEARGIK 134
|||||

RESULT 17
APJ_XENLA STANDARD; PRT; 353 AA.
ID APJ_XENLA
AC P79960; P70058;
DT 15-JUN-2002 (Rel. 41; Created)
DT 15-JUN-2002 (Rel. 41; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE G protein-coupled receptor APJ homolog (Angiotensin receptor related
protein) (Mesenchyme-associated serpentine receptor).
GN X-MR OR Xangi1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8353;
NM [1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE-97109511; PubMed-8951791;
RA Devic E., Paguereau L., Vernier P., Knibiehler B., Audigier Y.;
RT "Expression of a new G protein-coupled receptor X-mr is associated
with an endothelial lineage in Xenopus laevis.";
RN Mech. Dev. 59:129-140(1996).
[2]
RN RP SEQUENCE OF 1-303 FROM N.A.
RA Saha M.S., Oakes J.A., Miles R.R.;
RT "Xangi1, a novel Xenopus gene, is expressed in vascular precursor
cells.";
RN Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Putative receptor for an apelin-like peptide coupled to
G proteins that inhibit adenylate cyclase activity.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- TISSUE SPECIFICITY: At the gastrula stage, exclusively expressed
in the mesodermal layer and at the neurula stage, in the lateral
plate mesoderm. Larval expression is observed in the endothelium
of the primary blood vessels and the forming heart.
CC -I- DEVELOPMENTAL STAGE: First expressed at the late blastula stage,
increases during gastrulation and remains constant between neurula
and larva stages.
CC -I- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

```

```

CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X93045; CAA63612.1; -
DR EMBL: U72029; AAB17004.1; -
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSN.
DR PROSITE: PS00237; G_PROTEIN_REC_P1_1; 1.
DR PROSITE: PS00262; G_PROTEIN_REC_P1_2; 1.
KW G-protein coupled receptor; Glycoprotein; Transmembrane.
FT DOMAIN 1 38
FT TRANSMEM 39 61
FT DOMAIN 62 72
FT TRANSMEM 73 95
FT DOMAIN 96 109
FT TRANSMEM 110 131
FT DOMAIN 132 150
FT TRANSMEM 151 173
FT DOMAIN 174 206
FT TRANSMEM 207 229
FT DOMAIN 230 248
FT TRANSMEM 249 270
FT DOMAIN 271 294
FT TRANSMEM 295 317
FT DOMAIN 318 353
FT CARBOHYD 19 181
FT CARBOHYD 181 181
FT CONFLICT 150 150 P -> S (TN REF. 2).
SQ SEQUENCE 353 AA; 40303 MW; 1BF757D865057621 CRC64;

Query Match 3.3%; Score 7; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 IGLSLIL 175
DB 202 IGLSLIL 208
|||||

RESULT 18
OCD_AGRTS STANDARD; PRT; 354 AA.
ID OCD_AGRTS
AC P09773;
DT 01-MAR-1989 (Rel. 10; Created)
DT 15-JUN-2002 (Rel. 41; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Ornithine cyclodeaminase (EC 4.3.1.12) (OCD).
GN ARCB OR OCD OR AT06016 OR AGR_PTI_54.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Plasmid pTIC58.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
NCBI_TaxID=176299;
NM [1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE-88185308; PubMed-3281832;
RA Sans N., Schindler U., Schroeder J.;
RT "Ornithine cyclodeaminase from Ti plasmid C58: DNA sequence, enzyme
properties and regulation of activity by arginine.";
RN Eur. J. Biochem. 173:123-130(1988).
[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE-94321320; PubMed-8045881;
RA Zanker H., Lutz G., Langridge U., Langridge P., Kreusch D.,
RT Schroeder J.;
RT "Octopline and nopaline oxidases from Ti plasmids of Agrobacterium
tumefaciens: molecular analysis, relationship, and functional
characterization.";
RN J. Bacteriol. 176:4511-4517(1994).
[3]
RN RP SEQUENCE FROM N.A.
RX MEDLINE-21608550; PubMed-11743193;

```

Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 VAVSADP 19
111111
Db 59 VAVSADP 65

RESULT 14

YS23_CAEEL
ID YS23_CAEEL STANDARD; PRT; 208 AA.
AC 009365; 1-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 23.3 kDa protein ZK1320.3 in chromosome II.
GN ZK1320.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Berkis M.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: STRONG, TO C.ELEGANS ZK1320.2.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; 246934; GAA87041.1; -
DR WormPep: ZK1320.3; CE01701.
KW Hypothetical protein.
SQ SEQUENCE 208 AA: 23306 MW: 356A84E57A6C163B CRC64;

Query Match 3.3%; Score 7; DB 1; Length 208;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 MKFLIA 8
111111
Db 1 MKFLIA 7

RESULT 15
Y166_METUA
ID Y166_METUA STANDARD; PRT; 255 AA.
AC 05730; 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein M0106.
GN M0106.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weisslock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii";
RL Science 273:1058-1073(1996).
CC -1 SIMILARITY: BELONGS TO THE UPF0204 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U67473; AAB98148.1; -
DR TIGR; M0106; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 255 AA: 29072 MW: 6428DF7AEC802CE4 CRC64;

Query Match 3.3%; Score 7; DB 1; Length 255;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 MKFLIA 8
111111
Db 1 MKFLIA 7

RESULT 16

CHLO_RAT
ID CHLO_RAT STANDARD; PRT; 295 AA.
AC 003070; 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-chimaerin (Beta-chimerin).
GN CHN2 OR BCH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.

RC STRAIN=Missat; TISSUE=Testis;
RX MEDLINE=93179371; PubMed=8440677;
RA Leung T., How B.E., Manser E., Lim L.;
RA "Germ cell beta-chimaerin, a new GTPase-activating protein for
RT p12rac, is specifically expressed during the acrosomal assembly stage
RT in rat testis";
RL J. Biol. Chem. 268:3813-3816(1993).
CC -1 FUNCTION: GTPASE ACTIVATING PROTEIN FOR P21-RAC.
CC -1 SUBCELLULAR LOCATION: Membrane-associated (potential).
CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS: BETA-1 (SHOWN HERE) AND BETA-2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE CEREBELLAR BETA-2
CC FORM CONTAINS A SH2 DOMAIN.
CC -1 TISSUE SPECIFICITY: FOUND IN CEREBELLUM AND TESTIS.
CC -1 DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN LATE STAGE
CC SPERMATOCYTES. IN THE CEREBELLUM, EMERGENCE OF BETA-2 ISOFORM
CC COINCIDES WITH GRANULE CELLS MATURATION AND EXHIBITS POSTNATAL
CC DEVELOPMENTAL INCREASES. EXPRESSION IS SPECIFICALLY REDUCED IN
CC WEAVER MUTANT.
CC -1 SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

AC P52565;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Rho GDP-dissociation inhibitor 1 (Rho GDI 1) (Rho-GDI alpha).
 GN ARHGDI OR GDIA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94085490; PubMed=8262133;
 RA Leffers H., Nielsen M.S., Andersen A.H., Honore B., Madsen P.,
 RA Vandekerckhove J., Cells J.E.;
 RT "Identification of two human Rho GDP dissociation inhibitor proteins
 RT whose overexpression leads to disruption of the actin cytoskeleton.";
 RL Exp. Cell Res. 209:165-174(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Maeda A., Kaibuchi K., Takai Y.;
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Chuang T.H., Bokoch G.M.;
 RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Mulherson J.G., Schwinn D.A., Caron M.G., Liggett S.B.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RA Publ H.L. Iii, Ikeda S.R., Aronstam R.S.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Colon, Lung, Muscle, Skin, Tonsil, and Uterus;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 59-204.
 RX MEDLINE=97341226; PubMed=9195882;
 RA Keep N.H., Barnes M., Barsukov I., Badli R., Lian L.-Y., Segal A.W.,
 RA Moody P.C.E., Roberts G.C.K.;
 RT "A modulator of the family G proteins, rhogdi, binds these G proteins
 RT via an immunoglobulin-like domain and a flexible N-terminal arm.";
 RL Structure 5:623-633(1997).
 CC - FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF THE RHO
 CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
 CC SUBSEQUENT BINDING OF GTP TO THEM (BY SIMILARITY).
 CC - SUBUNIT: MONOMER (BY SIMILARITY).
 CC - CELLULAR LOCATION: Cytoplasmic.
 CC - SIMILARITY: BELONGS TO THE RHO GDI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X69550; CAA49281.1; -
 DR EMBL: D13989; BAA03096.1; -
 DR EMBL: M67579; AAA36566.1; -
 DR EMBL: X63863; CAA45344.1; -
 DR EMBL: AF498926; AAM21074.1; -
 DR EMBL: BC005851; AAH05851.1; -
 DR EMBL: BC005875; AAH05875.1; -
 DR EMBL: BC008701; AAH08701.1; -
 DR EMBL: BC009759; AAH09759.1; -
 DR EMBL: BC016031; AAH16031.1; -

DR EMBL: BC016185; AAH16185.1; -
 DR EMBL: BC024258; AAH24258.1; -
 DR EMBL: BC027730; AAH27730.1; -
 DR PDB: 1RHO; 15-OCT-97.
 DR Aarhus/Genet-2DPAGE; B118; IEF.
 DR Genew; HGNC:678; ARHGDI.
 DR MIM: 601925; -
 DR InterPro: IPR000406; Rho_GDI.
 DR Pfam: PF02115; Rho_GDI.1.
 DR PRINTS: PR00492; RHOGDI.
 KW GTPase activation; 3D-structure.
 FT CONFLICT 139 139 I -> V (IN REF. 3).
 SQ SEQUENCE 204 AA; 23207 MW; 59CB6FF42E3B3BCA CRC64;
 Query Match 3.3%; Score 7; DB 1; Length 204;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 VAASADP 19
 Db 59 VAASADP 65
 RESULT 13
 ID GDIR_MOUSE STANDARD; PRT; 204 AA.
 AC 099PT1; 099KC4;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Rho GDP-dissociation inhibitor 1 (Rho GDI 1) (Rho-GDI alpha) (GDI-1).
 GN ARHGDI OR GDII or C87222.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Minamitani T., Matsumoto K.;
 RT "Mouse cDNA sequence for Rhogdi-1.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF THE RHO
 CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
 CC SUBSEQUENT BINDING OF GTP TO THEM (BY SIMILARITY).
 CC - SUBUNIT: MONOMER (BY SIMILARITY).
 CC - CELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO THE RHO GDI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB055070; BAB21527.1; -
 DR EMBL: BC004732; AAH04732.1; -
 DR HSSP: P19803; 1GDF.
 DR PMMA-2DPAGE; 099PT1; -
 DR MGD: MGI:2144689; C87222.
 DR InterPro: IPR000406; Rho_GDI.
 DR Pfam: PF02115; Rho_GDI.1.
 DR PRINTS: PR00492; RHOGDI.
 KW GTPase activation.
 FT CONFLICT 55 55 L -> P (IN REF. 2).
 SQ SEQUENCE 204 AA; 23407 MW; 8ACB6FF44560842D8 CRC64;
 Query Match 3.3%; Score 7; DB 1; Length 204;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO00105; AAB91926.1; -
 DR InterPro: IPR004952; DUF269.
 DR Pfam: PF03270; DUF269.1
 KW Hypothetical protein; Nitrogen fixation; Plasmid.
 SQ SEQUENCE 162 AA; 18015 MW; 0B4D5C0FCB2BB61 CRC64;
 QY Query Match 3.3%; Score 7; DB 1; Length 162;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 144 LEISDEG 150
 100 LEISDEG 106
 Db 100 LEISDEG 106
 RESULT 10
 YE46_AQUAE STANDARD; PRT; 185 AA.
 ID YE46_AQUAE
 AC 067433;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AQ_1446.
 GN AQ_1446.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
 OC Aquifex.
 NC NCBITaxID=63363;
 RN NCBITaxID=63363;
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Anjaj M., Huder R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.";
 RL Nature 392:353-358(1998).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: STRONG, TO A.AEOLICUS AQ_1900.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO00741; AAC07402.1; -
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 1 21 POTENTIAL.
 SQ SEQUENCE 185 AA; 22670 MW; EDAL45E48ED739C9 CRC64;
 QY Query Match 3.3%; Score 7; DB 1; Length 185;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MMKFLI 7
 1 MMKFLI 7
 Db 1 MMKFLI 7
 RESULT 11

GDIR_BOVIN STANDARD; PRT; 204 AA.
 ID GDIR_BOVIN
 AC P19803;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Rho GDP-dissociation inhibitor 1 (Rho GDI 1) (Rho-GDI alpha).
 GN ARHGDI1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NC NCBITaxID=9913;
 RN NCBITaxID=9913;
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=91016432; PubMed=212066;
 RA Fukumoto Y., Kaibuchi K., Hori Y., Fujioke H., Araki S., Ueda T.,
 RA Kikuchi A., Takai Y.;
 RT "Molecular cloning and characterization of a novel type of regulatory
 RT protein (GDI) for the rho proteins, ras p21-like small GTP-binding
 RT proteins.";
 RL Oncogene 5:1321-1328(1990).
 CC -1-
 CC STRUCTURE BY NMR OF 60-204.
 RP MEDLINE=97337869; PubMed=9194563;
 RA Gossner Y.O., Nomanbhoy T.K., Aghazadeh B., Manor D., Combs C.,
 RA Cerione R.A., Rosen M.K.;
 RT "C-terminal binding domain of Rho GDP-dissociation inhibitor directs
 RT N-terminal inhibitory peptide to GTPases.";
 RL Nature 387:814-819(1997).
 CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF THE RHO
 CC PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
 CC SUBSEQUENT BINDING OF GTP TO THEM.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: BRAIN, LUNG, THYMUS, SPLEEN, SMALL INTESTINE,
 CC AND KIDNEY, AND WEAKLY IN HEART AND LIVER.
 CC -1- SIMILARITY: BELONGS TO THE RHO GDI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52689; CAA36916.1; -
 DR PIR: S12121; S12121.
 DR PDB: 1GDF; 19-NOV-97.
 DR PDB: 1AJW; 19-NOV-97.
 DR InterPro: IPR000406; Rho_GDI.
 DR Pfam: PF02115; Rho_GDI; 1.
 DR PRINTS: PR00492; RHOGDI.
 KW GTPase activation; Phosphorylation; 3D-structure.
 FT DOMAIN 66 83 HYDROPHOBIC.
 FT MOD_RES 101 101 PHOSPHORYLATION (BY PKA) (POTENTIAL).
 FT MOD_RES 115 115 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 SQ SEQUENCE 204 AA; 23421 MW; 49CE7DEB05D271CA CRC64;
 QY Query Match 3.3%; Score 7; DB 1; Length 204;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 13 VAWSADP 19
 59 VAWSADP 65
 Db 59 VAWSADP 65
 RESULT 12
 GDIR_HUMAN STANDARD; PRT; 204 AA.
 ID GDIR_HUMAN


```

FT MOD_RES 45 45 SIMILARITY).
SQ SEQUENCE 86 AA: 8805 MW: AE7B9CD819341A64 CRC64;
Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 DPAIAI 40
DB 70 DPAIAI 76

RESULT 7
GDIR_CAVPO
ID GDIR_CAVPO STANDARD: PRT: 111 AA.
AC P80237;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rho GDP-dissociation inhibitor 1 (Rho GDI 1) (Rho-GDI alpha)
DE (Fragments).
GN ARHGDI1A.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathii; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE.
RC STRAIN=Hartley; TISSUE=Macrophage;
RX MEDLINE=94039069; PubMed=8223583;
RA Pick E., Gorzalczany Y., Engel S.;
RT "Role of the rac1 p21-GDP-dissociation inhibitor for rho heterodimer
in the activation of the superoxide-forming NADPH oxidase of
macrophages."
RL Eur. J. Biochem. 217:441-455(1993).
CC -1- FUNCTION: REGULATES THE GDP/GTP EXCHANGE REACTION OF THE RHO
PROTEINS BY INHIBITING THE DISSOCIATION OF GDP FROM THEM, AND THE
SUBSEQUENT BINDING OF GTP TO THEM.
CC -1- FUNCTION: RAC1 P21/RHO GDI HETERODIMER IS THE ACTIVE COMPONENT OF
THE CYTOSOLIC FACTOR SIGMA 1, WHICH IS INVOLVED IN STIMULATION OF
THE NADPH OXIDASE ACTIVITY IN MACROPHAGES.
CC -1- SUBUNIT: FORMS A HETERODIMER WITH P21/RAC-1.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE RHO GDI FAMILY.
DR HSSP; P19803; IGDF.
KW GTPase activation.
FT NON_TER 1
FT NON_CONS 40 41
FT NON_CONS 46 47
FT NON_CONS 86 87
FT UNSURE 98 98
FT UNSURE 100 100
FT NON_TER 111
SQ SEQUENCE 111 AA: 12496 MW: 0FCCE35BEA40FE951 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 111;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSADP 19
DB 1 VAVSADP 7

RESULT 8
YCI3_ARCFU
ID YCI3_ARCFU STANDARD: PRT: 149 AA.
AC 029055;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Hypothetical protein AF1213.
GN AF1213.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001020; AAB90034.1; -
DR TIGR; AF1213; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 149 AA: 17778 MW: 955AE6E28093F77F CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 149;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 206 FKRELEK 212
DB 7 FKRELEK 13

RESULT 9
Y4XD_RHISN
ID Y4XD_RHISN STANDARD: PRT: 162 AA.
AC P55695;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 18.0 kDa protein Y4XD.
GN Y4XD.
OS Rhizobium sp. (strain NGR234).
OC plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes."
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: TO SIMILAR PROTEINS IN OTHER NITROGEN-FIXING BACTERIA.
CC THIS PROTEIN IS GENERALLY FOUND IN THE NIFX-NIFW INTERGENIC
REGION.
CC -1- SIMILARITY: STRONG, TO Y4VO.
CC -----

```

DR InterPro: IPR000840; Gag_MA.
 DR InterPro: IPR002079; Gag_p12.
 DR InterPro: IPR003036; Gag_p30.
 DR Pfam: PF01140; Gag_MA; 1.
 DR Pfam: PF01141; Gag_p12; 1.
 DR Pfam: PF02093; Gag_p30; 1.
 DR Core protein; Polyprotein; Myristate.
 KW PROPEP 1 77 LEADER PEPTIDE.
 FT CHAIN 78 204 CORE PROTEIN P15.
 FT CHAIN 205 274 CORE PROTEIN P12.
 FT CHAIN 275 522 CORE PROTEIN P30.
 FT CHAIN 523 536 CORE PROTEIN P10.
 FT LIPID 79 79 MYRISTATE.
 SQ SEQUENCE 536 AA; 60234 MW; F8BC80D612AC8702 CRC64;

Query Match 3.8%; Score 8; DB 1; Length 536;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203
 |||||
 DB 516 KEMTKVLA 523

RESULT 5
 GAG_FLV STANDARD; PRT; 580 AA.
 AC P10262; 085560; Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GAG polyprotein (contains: Core protein p15; Core protein p12; Core protein p30; Core protein p10).
 GN GAG.
 OS Feline leukemia virus.
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11768;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84216470; PubMed=6328019;
 RA Laprevotte I., Hampe A., Sherr C.J., Gilbert F.;
 RT Nucleotide sequence of the gag gene and gag-pol junction of feline leukemia virus.";
 RT J. Virol. 50:884-894(1984).
 CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: K01803; AAA43054.1; -
 DR EMBL: K01803; AAA43055.1; ALT_INIT.
 DR InterPro: IPR000840; Gag_MA.
 DR InterPro: IPR002079; Gag_p12.
 DR InterPro: IPR003036; Gag_p30.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00098; Zf-CCHC; 1.
 DR Pfam: PF01140; Gag_MA; 1.
 DR Pfam: PF01141; Gag_p12; 1.
 DR Pfam: PF02093; Gag_p30; 1.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS50158; ZF_CCHC; 1.
 KW Core protein; Polyprotein; Myristate; Zinc-finger.
 FT PROPEP 1 75 LEADER PEPTIDE.
 FT CHAIN 76 201 CORE PROTEIN P15.
 FT CHAIN 202 271 CORE PROTEIN P12.
 FT CHAIN 272 519 CORE PROTEIN P30.
 FT CHAIN 520 580 CORE PROTEIN P10.
 FT ZN_FING 547 564 CCHC-TYPE.

FT LIPID 76 76 MYRISTATE.
 SQ SEQUENCE 580 AA; 65195 MW; 10F8C3775B37042C CRC64;

Query Match 3.8%; Score 8; DB 1; Length 580;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203
 |||||
 DB 513 KEMTKVLA 520

RESULT 6
 PTHR_STRMU STANDARD; PRT; 86 AA.
 AC P45596;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phosphocarrier protein HPr (Histidine-containing protein).
 GN PTH.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NG5 / Serotype C;
 RX MEDLINE=94178918; PubMed=8132321;
 RA Boyd D.A., Cviltkovitch D.G., Hamilton I.R.;
 RT "Sequence and expression of the genes for HPr (ptsh) and enzyme I (ptsi) of the phosphoenolpyruvate-dependent phosphotransferase transport system from Streptococcus mutans.";
 RT Infect. Immun. 62:1156-1165(1994).
 RN [2]
 RP SEQUENCE.
 RC STRAIN=Ingbrilt;
 RX PubMed=8147873;
 RA Dasher S.G., Kirsbaum L., Hug N.L., Riley P.F., Reynolds E.C.;
 RT "Complete amino acid sequence and comparative molecular modelling of HPr from Streptococcus mutans Ingbrilt.";
 RT Biochem. Biophys. Res. Commun. 199:1297-1304(1994).
 CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE TRANSPORT SYSTEM. THE PHOSPHORYL GROUP FROM PHOSPHOENOLPYRUVATE (PEP) IS TRANSFERRED TO THE PHOSPHORYL CARRIER PROTEIN HPR BY ENZYME I. PHOSPHO-HPR THEN TRANSFERS IT TO THE PERKEASE (ENZYMES II/III). HPR IS COMMON TO ALL PTS.
 CC -1- ENZYME REGULATION: PHOSPHORYLATION ON SER-45 INHIBITS THE PHOSPHORYL TRANSFER FROM ENZYME I TO HPR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE HPR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L15191; AAA91092.1; -
 DR HSSP: P07515; IPTF.
 DR InterPro: IPR001020; HPr_Hisp_site.
 DR InterPro: IPR002114; HPr_Serp_site.
 DR InterPro: IPR000032; HPr_Protein.
 DR Pfam: PF00381; PTS-HPr; 1.
 DR TIGRFAMS: TIGR01003; PTS_HPr_family; 1.
 DR PROSITE; PS00369; PTS_HPR_HIS; 1.
 DR PROSITE; PS00589; PTS_HPR_SER; 1.
 KW Phosphotransferase system; Sugar transport; Phosphorylation.
 FT INIT_MET 0 0 PARTIAL.
 FT MOD_RES 14 14 PHOSPHORYLATION (BY ENZYME I) (BY

```

DB      181 VLSDLTALFODTVKREKMTKVLAPAFKRELERN 213
|||||
RESULT 2
ALL7_DERPT STANDARD: PRT; 215 AA.
ID      ALL7_DERPT STANDARD: PRT; 215 AA.
AC      P49273;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Mite allergen Der p 7 precursor (Der p VII).
GN      DERY7.
OS      Dermatophagoides pteronyssinus (House-dust mite).
OC      Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC      Acariformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;
OC      Dermatophagoides.
OX      NCBI_TaxID=69356;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Shen H.-D., Chua K.-Y., Lin K.-L., Hsieh K.-H., Thomas W.R.;
RT      "Molecular cloning of a house dust mite allergen with common antibody
RT      binding specificities with multiple components in mite extracts.";
RL      Clin. Exp. Allergy 23:934-940(1993).
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- SIMILARITY: BELONGS TO THE MITE GROUP 7 ALLERGEN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL: U37044; AAA80264.1; -
KM      Allergen; Signal.
FT      SIGNAL 1 17
FT      CHAIN 1 215 MITE ALLERGEN DER P 7.
FT      CARBOHD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 215 AA: 23861 MW: 63403DBSC8C1C0 CRC64:
Query Match 11.7%; Score 25; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 5.5e-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      155 TSFEVROFANVYVNHIGLSILDPIT 179
DB      155 TSFEVROFANVYVNHIGLSILDPIT 179
|||||
RESULT 3
DHON_METGL STANDARD: PRT; 412 AA.
ID      DHON_METGL STANDARD: PRT; 412 AA.
AC      P37144;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Homoserine dehydrogenase (EC 1.1.1.3) (HDH).
GN      HOM.
OS      Methylobacillus glycogenes.
OC      Bacteria; Proteobacteria; beta subdivision; Methylophilus group;
OC      Methylobacillus.
OX      NCBI_TaxID=406;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=ATCC 213171;
RA      MEDLINE=94161493; PubMed=81117070;
RA      McCoyama H., Maki K., Anazawa H., Ishino S., Teshiba S.;
RT      "Cloning and nucleotide sequences of the homoserine dehydrogenase
RT      genes (hdm) and the threonine synthase genes (thrc) of the Gram-
RT      negative obligate methylotroph Methylobacillus glycogenes.";
RL      Appl. Environ. Microbiol. 60:1111-1119(1994).

```

```

CC      -1- CATALYTIC ACTIVITY: L-homoserine + NAD(P)(+) = L-aspartate 4-
CC      semialdehyde + NAD(P)H.
CC      -1- PATHWAY: THIRD STEP IN CONVERSION OF L-ASPARTATE TO HOMOSERINE.
CC      HOMOSERINE PARTICIPATES IN THE BIOSYNTHESIS OF THEONINE AND THEN
CC      ISOLEUCINE AND IN THE BIOSYNTHESIS OF METHIONINE.
CC      -1- SIMILARITY: BELONGS TO THE HOMOSERINE DEHYDROGENASE FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL: D14071; BAA40415.1; -
DR      InterPro: IPR002912; ACT.
DR      InterPro: IPR001342; Homoserine_dh.
DR      InterPro: IPR005106; NAD_binding_3.
DR      Pfam: PF00742; Homoserine_dh; 1.
DR      Pfam: PF01842; ACT; 1.
DR      Pfam: PF03447; NAD_binding_3; 1.
DR      PROSITE: PS01042; HOMOSER-DHGENASE; 1.
KM      Oxidoreductase: NADP; Threonine biosynthesis: Isoleucine biosynthesis;
KM      Methionine biosynthesis.
FT      NP_BIND 9 16 NADP (POTENTIAL).
SQ      SEQUENCE 412 AA: 44818 MW: 613A1B7FDECEFA4A CRC64:
OY      34 DDALIAIE 41
DB      386 DDALIAIE 393
|||||
RESULT 4
GAG_FSVMD STANDARD: PRT; 536 AA.
ID      GAG_FSVMD STANDARD: PRT; 536 AA.
AC      P03340;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      GAG polypeptide [contains: Core protein p15; Core protein p12; Core
DE      protein p30; Core protein p10].
GN      GAG.
OS      Feline sarcoma virus (strain McDonough).
OC      Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX      NCBI_TaxID=11778;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=84119469; PubMed=6582485;
RA      Hampe A., Gobet M., Sherr C.J., Galibert F.;
RT      "Nucleotide sequence of the feline retroviral oncogene v-fms shows
RT      unexpected homology with oncogenes encoding tyrosine-specific protein
RT      kinases.";
RL      Proc. Natl. Acad. Sci. U.S.A. 81:85-89(1984).
CC      -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-FMS
CC      POLYPEPTIDE.
CC      -1- SIMILARITY: TO MOUSE TESTOSTERONE-REGULATED RP2 PROTEIN.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL: K01643; AAA43045.1; ALF_TERM.
DR      PIR: A03938; F0WVMD.

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:10:10 ; Search time 13 Seconds

(without alignments)
679.574 Million cell updates/sec

Title: US-10-024-955-7

Percent score: 213
Sequence: 1 MMKFLIAAFAVAVSADPT.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Swisprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	213	1 ALL7_DERFA	Q26456 dermatophag
2	25	11.7	215	1 ALL7_DERPT	P45273 dermatophag
3	8	3.8	412	1 DHON_METCL	P37144 methylolaci
4	8	3.8	536	1 GAG_FSVMD	P03340 feline sarc
5	8	3.8	580	1 GAG_FLV	P10262 feline leuk
6	7	3.3	86	1 PTHP_STRMU	P45596 streptococc
7	7	3.3	111	1 GDIR_CAVPO	P80237 cavia porce
8	7	3.3	149	1 YC13_ARCFU	O29055 archaeoglob
9	7	3.3	162	1 Y4XD_RHISN	P35695 rhizobium s
10	7	3.3	185	1 YE46_AQDAE	O67403 bos taurus
11	7	3.3	204	1 GDIR_BOVIN	P19803 bovis aequi
12	7	3.3	204	1 GDIR_HUMAN	P52565 homo sapien
13	7	3.3	204	1 GDIR_MOUSE	O99P11 mus musculu
14	7	3.3	208	1 VS23_CAEBL	Q03365 caenorhabdi
15	7	3.3	255	1 Y166_METVA	Q57630 methanococc
16	7	3.3	295	1 CHIO_RAT	O03070 rattus norv
17	7	3.3	353	1 APJ_XENLA	P79960 xenopus lae
18	7	3.3	354	1 OCD_AGR5	P09773 agrobacteri
19	7	3.3	356	1 OCD_AGR4	O59701 agrobacteri
20	7	3.3	434	1 Y166_CAEBL	Q11082 caenorhabdi
21	7	3.3	460	1 ORC5_DROME	Q24169 drosophila
22	7	3.3	468	1 CHIO_HUMAN	P52757 homo sapien
23	7	3.3	500	1 ARAA_YERPE	P58540 yersinia pe
24	7	3.3	511	1 CADL_DROME	O9127 drosophila
25	7	3.3	556	1 CG48_SCHHO	P78750 schizosacch
26	7	3.3	833	1 VIRA_AGR5	P18540 agrobacteri
27	7	3.3	886	1 SYA_BARBA	P70865 bartonella
28	7	3.3	2009	1 SEC7_YEAST	P11075 saccharomyc
29	6	2.8	40	1 LHBI_ECTHL	P80106 ectothiorho
30	6	2.8	48	1 LHBI_RHOTE	P80590 rhodocyclu
31	6	2.8	48	1 LHBI_RHOTE	P80591 rhodocyclu
32	6	2.8	52	1 A95E_DROME	P16548 drosophila
33	6	2.8	77	1 ACP_BACHD	O9ka04 bacillus ha

34	6	2.8	82	1 ATPH_ANTSP	Q02851 antithamnio
35	6	2.8	82	1 ATPH_CYACA	O9em30 cyanidium c
36	6	2.8	82	1 ATPH_PORU	P51246 porphyra pu
37	6	2.8	83	1 ATPH_GALSU	P35013 galdieria s
38	6	2.8	83	1 ATPH_PAVLU	P28530 pavlova lut
39	6	2.8	86	1 YK45_PSEAE	O91270 pseudomonas
40	6	2.8	88	1 PTHP_LACCA	O9kfjv3 lactobacill
41	6	2.8	94	1 ES61_MYCTU	P96364 mycobacteri
42	6	2.8	101	1 NUDC_PLEBO	O00244 plectonema
43	6	2.8	117	1 Y13K_BP74	P39504 bacterioph
44	6	2.8	124	1 GCSH_THEMA	O9wy55 thermotoga
45	6	2.8	135	1 JANA_DROME	P20348 drosophila

ALIGNMENTS

RESULT 1	ID	ALL7_DERFA	STANDARD:	PRT:	213 AA.
AC	Q26456:				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Mite allergen Der f 7 precursor (Der f VII).				
GN	DERF7.				
OS	Dermatophagoides farinae (House-dust mite).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;				
OC	Acariformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;				
OX	Dermatophagoides.				
OX	NCBI_TaxID=6954;				
RP	[1]				
RX	SEQUENCE FROM N.A.				
RA	MEDLINE=96120794; PubMed=8556554;				
RA	Shen H.-D., Chua K.-Y., Lin W.-L., Hsieh K.-H., Thomas W.R.;				
RT	"Molecular cloning and immunological characterization of the house				
RT	dust mite allergen Der f 7.";				
RL	Clin. Exp. Allergy 25:1000-1006(1995).				
CC	-1- SUBCELLULAR LOCATION: Secreted.				
CC	-1- SIMILARITY: BELONGS TO THE MITE GROUP 7 ALLERGEN FAMILY.				
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; S80655; AAB35977.1; -				
KM	Allergen; Signal.				
FT	SIGNAL	1	17	POTENTIAL.	
FT	CHAIN	18	213	MITE ALLERGEN DER F 7.	
FT	CARBOHD	151	151	N-LINKED (GLCNAG. . .) (POTENTIAL).	
SO	SEQUENCE	213 AA; 23627 MW; 3CF1F529107B7808 CRC64;			
Query Match	Best Local Similarity	100.0%; Score 213; DB 1; Length 213;			
Matches	213; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY	1 MMKFLIAAFAVAVSADPTIHYDKITTEINKAIDAIAIEQSEITIDPKVDPHADKFER 60				
DB	1 MMKFLIAAFAVAVSADPTIHYDKITTEINKAIDAIAIEQSEITIDPKVDPHADKFER 60				
DB	61 HVGIVDFGELAMRNIEARGLKQMRQGDANVKGEGIVKAHLILGVHDIYSMEYDAY 120				
OY	61 HVGIVDFGELAMRNIEARGLKQMRQGDANVKGEGIVKAHLILGVHDIYSMEYDAY 120				
OY	121 KIGDHPPTHTVSDIODPVVALSLTSDSGNTTMSFEVROANVNNHTGSLIDPTFG 180				
DB	121 KIGDHPPTHTVSDIODPVVALSLTSDSGNTTMSFEVROANVNNHTGSLIDPTFG 180				
OY	181 VLSVLTATIFQPTVRKEMTKVLAPAFKRELEKN 213				

```
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4892
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-4892
```

```
Query Match          2.8%; Score 6; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 121 KLGDLH 126
    |||||
Db 154 KLGDLH 159
```

```
RESULT 50
US-09-815-242-10483
; Sequence 10483, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
```

```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10483
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10483
```

```
Query Match          2.8%; Score 6; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 121 KLGDLH 126
    |||||
Db 154 KLGDLH 159
```

```
Search completed: February 21, 2003, 10:23:01
Job time : 17 secs
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (51)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (123)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (132)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (201)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (202)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-884-32

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 281;
Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 GELAMR 74
      |||||
Db 150 GELAMR 155

RESULT 46
US-09-925-297-803
; Sequence 803, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 803
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (225)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-803

Query Match
Best Local Similarity 100.0%; Score 6; DB 10; Length 281;
Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 GELAMR 74
      |||||
Db 173 GELAMR 178

RESULT 47
US-09-971-536-67
; Sequence 67, Application US/09971536
```

```
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christenson; Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/N201/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-67

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 282;
Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATIA 39
      |||||
Db 212 DDATIA 217

RESULT 48
US-09-764-870-292
; Sequence 292, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 292
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-292

Query Match
Best Local Similarity 100.0%; Score 6; DB 10; Length 293;
Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 VLSDVL 186
      |||||
Db 232 VLSDVL 237

RESULT 49
US-09-815-242-4892
; Sequence 4892, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
```

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentln ver. 3.0
SEQ ID NO 4408
LENGTH: 251
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4408

Query Match 2.8%; Score 6; DB 9; Length 251;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 148 DEGNIT 153
Db 54 DEGNIT 59

RESULT 43
US-09-764-884-23
Sequence 23, Application US/09764884
Patent No. US20020161208A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT203
CURRENT APPLICATION NUMBER: US/09/764,884
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 23
LENGTH: 257
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-884-23

Query Match 2.8%; Score 6; DB 9; Length 257;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 GELAMR 74
Db 149 GELAMR 154

RESULT 44
US-09-864-761-46564
Sequence 46564, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-x-1

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 46564
LENGTH: 277
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004614.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
OTHER INFORMATION: SWISSPROT HIT: P35446, EVALU 3.00e-12
OTHER INFORMATION: EST_HUMAN HIT: AW612526.1, EVALU 6.00e-65
US-09-864-761-46564

Query Match 2.8%; Score 6; DB 10; Length 277;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 KGEEGI 98
Db 84 KGEEGI 89

RESULT 45
US-09-764-884-32
Sequence 32, Application US/09764884
Patent No. US20020161208A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT203
CURRENT APPLICATION NUMBER: US/09/764,884
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 38

```
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; FILE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 231
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
; FEATURE:
; OTHER INFORMATION: putative ORF from GenBank #U32720
US-09-816-028A-49
```

```
Query Match          2.8%: Score 6; DB 10; Length 231;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 113 SMEYDL 118
    |||||
Db 128 SMEYDL 133
```

```
RESULT 39
US-09-953-349-8
; Sequence 8, Application US/09953349
; Patent No. US20020099014A1
; GENERAL INFORMATION:
; APPLICANT: Brennan, Miles
; TITLE OF INVENTION: Method for Treatment of Insulin Resistance in Obesity and Diabetes
; FILE REFERENCE: 3718-7
; CURRENT APPLICATION NUMBER: US/09/953,349
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: 60/232,292
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 235
; TYPE: PRF
; ORGANISM: Mus musculus
US-09-953-349-8
```

```
Query Match          2.8%: Score 6; DB 10; Length 235;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 206 FKRELE 211
    |||||
Db 162 FKRELE 167
```

```
RESULT 40
US-09-738-626-4458
; Sequence 4458, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
```

```
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4458
; LENGTH: 237
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4458
```

```
Query Match          2.8%: Score 6; DB 9; Length 237;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 VAFVAV 15
    |||||
Db 23 VAFVAV 28
```

```
RESULT 41
US-09-767-041-11
; Sequence 11, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 244
; TYPE: PRF
; ORGANISM: Streptococcus suis
; NAME/KEY: misc-feature
; OTHER INFORMATION: ORF2X
US-09-767-041-11
```

```
Query Match          2.8%: Score 6; DB 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 133 SDIODF 138
    |||||
Db 214 SDIODF 219
```

```
RESULT 42
US-09-738-626-4408
; Sequence 4408, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
```



```
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10020
; LENGTH: 201
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-815-242-10020
```

```
Query Match          2.8%; Score 6; DB 10; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 34 DDAIAA 39
    |||||
Db 186 DDAIAA 191
```

```
RESULT 35
US-09-924-256A-44
; Sequence 44, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Yiyuan
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 214
; TYPE: PRF
; ORGANISM: Parmelia sulcata
US-09-924-256A-44
```

```
Query Match          2.8%; Score 6; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 34 DDAIAA 39
    |||||
Db 157 DDAIAA 162
```

```
RESULT 36
US-09-143-127-7
; Sequence 7, Application US/09143127
; Patent No. US20020136742A1
; GENERAL INFORMATION:
; APPLICANT: Kousoulas, K.
; APPLICANT: Chouljenko, V.
; APPLICANT: Baghian, A.
; APPLICANT: Tully, Jr., T.
; TITLE OF INVENTION: Vaccines for Chlamydia psittaci
; FILE REFERENCE: 21099.0056
; CURRENT APPLICATION NUMBER: US/09/143,127
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: 60/057,147
; EARLIER FILING DATE: 1997-08-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 215
; TYPE: PRF
; ORGANISM: Chlamydia psittaci (cockatiel)
US-09-143-127-7
```

```
Query Match          2.8%; Score 6; DB 10; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 181 VLSDVL 186
    |||||
Db 158 VLSDVL 163
```

```
RESULT 37
US-09-143-127-1
; Sequence 1, Application US/09143127
; Patent No. US20020136742A1
; GENERAL INFORMATION:
; APPLICANT: Kousoulas, K.
; APPLICANT: Chouljenko, V.
; APPLICANT: Baghian, A.
; APPLICANT: Tully, Jr., T.
; TITLE OF INVENTION: Vaccines for Chlamydia psittaci
; FILE REFERENCE: 21099.0056
; CURRENT APPLICATION NUMBER: US/09/143,127
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: 60/057,147
; EARLIER FILING DATE: 1997-08-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 222
; TYPE: PRF
; ORGANISM: Chlamydia psittaci (cockatiel)
US-09-143-127-1
```

```
Query Match          2.8%; Score 6; DB 10; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 181 VLSDVL 186
    |||||
Db 165 VLSDVL 170
```

```
RESULT 38
US-09-816-028A-49
; Sequence 49, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
US-09-816-028A-49
```

; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17

; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,065
; PRIOR FILING DATE: 2000-09-14

Query Match

Best Local Similarity 2.8%; Score 6; DB 9; Length 199;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LSLEIS 147
DB 147 LSLEIS 152

RESULT 34

US-09-815-242-10020
; Sequence 10020, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel

RESULT 31
US-09-839-894-34
; Sequence 34, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: USFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-34

Query Match 2.8%; Score 6; DB 9; Length 168;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 FVAVSA 17
|11111|
Db 16 FVAVSA 21

RESULT 32
US-09-839-894-37
; Sequence 37, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: USFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-37

Query Match 2.8%; Score 6; DB 9; Length 168;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 FVAVSA 17
|11111|
Db 16 FVAVSA 21

RESULT 33
US-09-989-442-135

; Sequence 135, Application US/09989442
; Publication No. US20030013649A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ208
; CURRENT APPLICATION NUMBER: US/09/989,442
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868

RESULT 27
US-09-925-299-1210
; Sequence 1210, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1210
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (128)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (129)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1210

Query Match 2.8%; Score 6; DB 10; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 HVGIVD 66
DB 22 HVGIVD 27

RESULT 28
US-09-823-038A-35
; Sequence 35, Application US/09823038A
; Patent No. US2002005835A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murlison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Mouse
US-09-823-038A-35

Query Match 2.8%; Score 6; DB 10; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MKFLI 7

DB 1 MKFLI 6
RESULT 29
US-09-792-246-4
; Sequence 4, Application US/09792246
; Patent No. US20020058018A1
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; TITLE OF INVENTION: A No. US20020058018a1 Interleukin-3 and Uses Thereof
; FILE REFERENCE: 28110/35850
; CURRENT APPLICATION NUMBER: US/09/792,246
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 09/376,732
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 09/024,820
; PRIOR FILING DATE: 1998-02-17
; PRIOR APPLICATION NUMBER: US 09/177,467
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: US 09/209,534
; PRIOR FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-792-246-4

Query Match 2.8%; Score 6; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 LSLEIS 147
DB 147 LSLEIS 152

RESULT 30
US-10-001-876-197
; Sequence 197, Application US/10001876
; Patent No. US20020177140A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes a
; FILE REFERENCE: DEX-0285
; CURRENT APPLICATION NUMBER: US/10/001,876
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,186
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 197
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-876-197

Query Match 2.8%; Score 6; DB 9; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 FANVNV 167
DB 75 FANVNV 80

US-09-793-306-27
; Sequence 27, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 27
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Tb475 ORF-2
; NAME/KEY: MOD.RES
; LOCATION: (20)
; OTHER INFORMATION: Xaa = any amino acid
US-09-793-306-27

Query Match 2.8%; Score 6; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 183 SDVLT A 188
| | | | |
Db 33 SDVLT A 38

RESULT 24
US-09-793-306-29
; Sequence 29, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 29
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Tb488 ORF-2
US-09-793-306-29

Query Match 2.8%; Score 6; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 183 SDVLT A 188
| | | | |
Db 33 SDVLT A 38

RESULT 25
US-09-823-038A-34
; Sequence 34, Application US/09823038A
; Patent No. US2002005835A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevil
; APPLICANT: Orrust, Rene
; APPLICANT: Kumbie, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Mouse
US-09-823-038A-34

Query Match 2.8%; Score 6; DB 10; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 MKFLLI 7
| | | | |
Db 1 MKFLLI 6

RESULT 26
US-09-975-719-150
; Sequence 150, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-150

Query Match 2.8%; Score 6; DB 9; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 170 GGLSIL 175
| | | | |
Db 23 GGLSIL 28

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
US-09-073-009-27

Query Match 2.8%; Score 6; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SDVLT 188
Db 33 SDVLT 38

RESULT 18
US-09-073-009-29
Sequence 29, Application US/09073009
Patent No. US20010012888A1
GENERAL INFORMATION:
APPLICANT: Alderson, Mark
APPLICANT: Dillon, David C.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
US-09-073-009-29

Query Match 2.8%; Score 6; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SDVLT 188
Db 33 SDVLT 38

RESULT 19
US-09-023-588-25
Sequence 25, Application US/09023588
Patent No. US20020081579A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Alderson, Mark R.
TITLE OF INVENTION: METHOD FOR THE ISOLATION OF NOVEL ANTIGENS
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,588
FILING DATE: 14-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.445
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
US-09-023-588-25

Query Match 2.8%; Score 6; DB 10; Length 94;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 SDVLT 188
Db 33 SDVLT 38

RESULT 20
US-09-023-588-27
Sequence 27, Application US/09023588

RESULT 14
US-09-971-536-45
; Sequence 45, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; TITLE OF INVENTION: Using them
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 88
; TYPE: PRF
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-45

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 88;
Pred. No. 94;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 DATAI 40
Db 72 DATAI 77

RESULT 15
US-09-950-933A-55
; Sequence 55, Application US/09950933A
; Patent No. US2002016141A1
; GENERAL INFORMATION:
; APPLICANT: Simmons, Carl R.
; APPLICANT: Navarro, Pedro
; TITLE OF INVENTION: Antimicrobial Peptides and Methods of
; TITLE OF INVENTION: Use
; FILE REFERENCE: 35718/238472
; CURRENT APPLICATION NUMBER: US/09/950,933A
; CURRENT FILING DATE: 2001-09-11
; PRIOR APPLICATION NUMBER: 60/232,569
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55
; LENGTH: 88
; TYPE: PRF
; ORGANISM: Glycine max
US-09-950-933A-55

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 88;
Pred. No. 94;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VAFVAV 15
Db 3 VAFVAV 8

RESULT 16
US-09-073-009-25
; Sequence 25, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
US-09-073-009-25

Query Match
Best Local Similarity 100.0%; Score 6; DB 10; Length 94;
Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
Db 33 SDVLTA 38

RESULT 17
US-09-073-009-27
; Sequence 27, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; TUBERCULOSIS AND


```
? LENGTH: 43
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AL15773.3
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.4
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.8
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.6
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.8
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.4
? US-09-864-761-40280

Query Match
Best Local Similarity 2.8%; Score 6; DB 10; Length 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 LIAVA 11
Db 5 LIAVA 10

RESULT 12
? US-09-864-761-35621
? Sequence 35621, Application US/09864761
? Patent No. US20020048763A1
? GENERAL INFORMATION:
? APPLICANT: Penn, Sharon G.
? APPLICANT: Rank, David R.
? APPLICANT: Hanzel, David K.
? APPLICANT: Chen, Wensheng
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
? TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
? FILE REFERENCE: Aemica X-1
? CURRENT APPLICATION NUMBER: US/09/864,761
? PRIOR FILING DATE: 2001-05-23
? PRIOR APPLICATION NUMBER: US 60/180,312
? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: US 60/207,456
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: US 09/632,366
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: GB 24263.6
? PRIOR FILING DATE: 2000-10-04
? PRIOR APPLICATION NUMBER: US 60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
```

```
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 35621
? LENGTH: 58
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC007089.2
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
? OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
? OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8
? OTHER INFORMATION: EXPRESSED IN EST_HUMAN HIT: BF129244.1, EVALUOE 1.20e+00
? US-09-864-761-35621

Query Match
Best Local Similarity 2.8%; Score 6; DB 10; Length 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AVSADP 19
Db 39 AVSADP 44

RESULT 13
? US-09-738-626-6222
? Sequence 6222, Application US/09738626
? Publication No. US20020197605A1
? GENERAL INFORMATION:
? APPLICANT: NAKAGAWA, SATOSHI
? APPLICANT: MIZOGUCHI, HIROSHI
? APPLICANT: ANDO, SEIKO
? APPLICANT: HAYASHI, MIKIRO
? APPLICANT: OCHIAI, KEIKO
? APPLICANT: YOKOI, HARUHIKO
? APPLICANT: TATEISHI, NAOKO
? APPLICANT: SENOH, AKIHIRO
? APPLICANT: IKEDA, MASATO
? APPLICANT: OZAKI, AKIO
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-125
? CURRENT APPLICATION NUMBER: US/09/738,626
? CURRENT FILING DATE: 2000-12-18
? PRIOR APPLICATION NUMBER: JP 99/377484
? PRIOR FILING DATE: 1999-12-16
? PRIOR APPLICATION NUMBER: JP 00/159162
? PRIOR FILING DATE: 2000-04-07
? PRIOR APPLICATION NUMBER: JP 00/280988
? PRIOR FILING DATE: 2000-08-03
? NUMBER OF SEQ ID NOS: 7059
? SOFTWARE: PatentIn ver. 3.0
? SEQ ID NO 6222
? LENGTH: 69
? TYPE: PRT
? ORGANISM: Corynebacterium glutamicum
? US-09-738-626-6222

Query Match
Best Local Similarity 2.8%; Score 6; DB 9; Length 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 FLIIAA 9
Db 17 FLIIAA 22
```

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 9.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EST_HUMAN HIT: AW874553.1, EVALUE 2.00e-04
US-09-864-761-36329

Query Match
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 TKVLAP 204
Db 10 TKVLAP 15

RESULT 10
US-09-864-761-41249
Sequence 41249, Application US/09864761
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 41249
LENGTH: 33
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC012308.4

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.82
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EST_HUMAN HIT: BF082286.1, EVALUE 8.00e-12
OTHER INFORMATION: SWISSPROT HIT: P07664, EVALUE 6.50e+00
US-09-864-761-41249

Query Match
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 ITTEIN 30
Db 12 ITTEIN 17

RESULT 11
US-09-864-761-40280
Sequence 40280, Application US/09864761
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 40280

RESULT 7
US-09-793-306-73
; Sequence 73, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 73
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:peptide
US-09-793-306-73

Query Match 2.8%; Score 6; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 183 SDVLT A 188
| | | | |
Db 8 SDVLT A 13

RESULT 8
US-09-793-306-75
; Sequence 75, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:peptide
US-09-793-306-75

Query Match 2.8%; Score 6; DB 10; Length 15;

Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLT A 188
| | | | |
Db 3 SDVLT A 8

RESULT 9
US-09-864-761-36329
; Sequence 36329, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36329
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008015.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5

```
;
; COMPUTER: IBM PC
; OPERATING SYSTEM: Windows 98
; SOFTWARE: Wordperfect 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,955C
; FILING DATE: 24-September-2001
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/938,275
; FILING DATE: 22-August-2001
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dwyer, Patrick M.
; REGISTRATION NUMBER: 32,411
; REFERENCE/DOCKET NUMBER: PROTEO.P03CI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 343-7074
; TELEFAX: (206) 343-7085
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; ORIGINAL SOURCE:
; ORGANISM: MOUSE
; FEATURE:
; OTHER INFORMATION: Also referred to in the specification as "LAM-L"
US-09-962-955C-20

Query Match          2.8%; Score 6; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VANSAD 18
Db 6 VAVSAD 11

RESULT 5
US-09-073-009-73
; Sequence 73, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
;
; TUBERCULOSIS AND ME
```

```
;
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
US-09-073-009-73

Query Match          2.8%; Score 6; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
Db 8 SDVLTA 13

RESULT 6
US-09-073-009-75
; Sequence 75, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
US-09-073-009-75

Query Match          2.8%; Score 6; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
Db 3 SDVLTA 8

TUBERCULOSIS AND
```

```

;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-024-955-7

```

Query Match	100.0%	Score 213	DB 9	Length 213
Best Local Similarity	100.0%	Pred. No. 1e-200		
Matches 213	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	MMKELLIAAVFVAVNSAP	IHYDKITEEINKAIDDAIAIQSE	ETIDPMKVPBHADEF	60
Db	1	MMKELLIAAVFVAVNSAP	IHYDKITEEINKAIDDAIAIQSE	ETIDPMKVPBHADEF	60
QY	61	HVGIVDFEKGELAMRNI	EARGLKQKROGDANV	KGEEGIVAKHLIGVHDDIVSMEYD	120
Db	61	HVGIVDFEKGELAMRNI	EARGLKQKROGDANV	KGEEGIVAKHLIGVHDDIVSMEYD	120
QY	121	KLGLDHPHTTHVYISDI	QIDPVALLSLEISDEG	NIYMTSEFVROFANVNHIGLSILDP	180
Db	121	KLGLDHPHTTHVYISDI	QIDPVALLSLEISDEG	NIYMTSEFVROFANVNHIGLSILDP	180
QY	181	VLSDLVLTAIPOD	YVKEKTKVLAPAFKRELE	ELKN 213	
Db	181	VLSDLVLTAIPOD	YVKEKTKVLAPAFKRELE	ELKN 213	

RESULT 2
US-10-024-955-2

```

: TITLE OF INVENTION: Allergenic Proteins and Peptides From
:                               House Dust Mite and Uses Therefor
:
: NUMBER OF SEQUENCES: 15
:
: CORRESPONDENCE ADDRESS:
:
:

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-033CP
TELECOMMUNICATION INFORMATION:

```

? INFORMATION FOR SEQ ID NO: 2:
?     SEQUENCE CHARACTERISTICS:
?         LENGTH: 215 amino acids
?         TYPE: amino acid
?         TOPOLOGY: linear
?     MOLECULE TYPE: protein
?     SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-024-955-2

```

Query Match	11.7%;	Score 25;	DB 9;	Length 215;
Best Local Similarity	100.0%;	Pred. No.	5.9e-17;	

	Matches	25;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0.
QY	155	TSEFVRQFANVYNHIGLSILDPF	179							
Db	155	TSEFVRQFANVYNHIGLSILDPF	179							

RESULT 3
US-09-815-242-13280
; Sequence 13280, Application US/09815242
; Patent No. US20020061569A1

```
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; TITLE OF INVENTION: Prokaryotes
```

```

; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13280

```

Query Match	3.3%	Score 7;	DB 10;	Length 87;
Best Local	Similarity 100.0%;	Pred. No. 9.8;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	34	DDAIAAI	40
		IIIIII	
Db	71	DDAIAAI	77

RESULT 4
US-09-962-955C-20
: Sequence 20, Application US/09962955C
: Publication NO. US20030013648A1
: GENERAL INFORMATION:
: APPLICANT: Gerardo M. Castillo
: APPLICANT: Alan D. Snow
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Patrick M. Dwyer
: STREET: Proteotech, Inc, 1818 Westlake Avenue N, Suite 1100
: CITY: Seattle
: STATE: WA (Washington)
: COUNTRY: United States of America
: ZIP: 98109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette - 3.50 inch, 1.44 Mb storage

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:15:21 ; Search time 15 Seconds
(without alignments)
441.191 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 213
Sequence: 1 MMKFLIAAVFAVAVSADPI.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 156504 seqs, 31069816 residues

Word size : 0

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	213	9	US-10-024-955-7
2	25	11.7	215	9	US-10-024-955-2
3	7	3.3	87	10	US-09-815-242-13280
4	6	2.8	12	9	US-09-962-955C-20
5	6	2.8	15	10	US-09-073-009-73
6	6	2.8	15	10	US-09-073-009-73
7	6	2.8	15	10	US-09-793-306-75
8	6	2.8	15	10	US-09-793-306-75
9	6	2.8	21	10	US-09-864-761-36329
10	6	2.8	33	10	US-09-864-761-41249
11	6	2.8	43	10	US-09-864-761-40280
12	6	2.8	58	10	US-09-864-761-35621
13	6	2.8	6	9	US-09-738-626-6222
14	6	2.8	88	9	US-09-971-536-45
15	6	2.8	88	9	US-09-950-933A-55
16	6	2.8	94	10	US-09-073-009-25
17	6	2.8	94	10	US-09-073-009-27
18	6	2.8	94	10	US-09-073-009-29
19	6	2.8	94	10	US-09-023-588-25

20	6	2.8	94	10	US-09-023-588-27	Sequence 27, Appl
21	6	2.8	94	10	US-09-023-588-29	Sequence 29, Appl
22	6	2.8	94	10	US-09-793-306-25	Sequence 25, Appl
23	6	2.8	94	10	US-09-793-306-27	Sequence 27, Appl
24	6	2.8	94	10	US-09-793-306-29	Sequence 29, Appl
25	6	2.8	102	10	US-09-823-038A-34	Sequence 34, Appl
26	6	2.8	128	9	US-09-975-719-150	Sequence 150, App
27	6	2.8	129	10	US-09-925-299-1210	Sequence 1210, Ap
28	6	2.8	147	10	US-09-823-038A-45	Sequence 35, Appl
29	6	2.8	152	10	US-09-792-026-4	Sequence 45, Appl
30	6	2.8	161	9	US-10-001-876-197	Sequence 197, App
31	6	2.8	168	9	US-09-839-894-34	Sequence 34, Appl
32	6	2.8	168	9	US-09-839-894-37	Sequence 37, Appl
33	6	2.8	199	9	US-09-989-442-135	Sequence 135, App
34	6	2.8	201	10	US-09-815-242-10020	Sequence 10020, A
35	6	2.8	214	10	US-09-924-256A-44	Sequence 44, Appl
36	6	2.8	215	10	US-09-143-127-7	Sequence 7, Appl
37	6	2.8	222	10	US-09-143-127-1	Sequence 1, Appl
38	6	2.8	231	10	US-09-816-028A-49	Sequence 49, Appl
39	6	2.8	235	10	US-09-953-349-8	Sequence 8, Appl
40	6	2.8	237	9	US-09-738-626-4458	Sequence 4458, Ap
41	6	2.8	244	10	US-09-767-041-11	Sequence 11, Appl
42	6	2.8	251	9	US-09-738-626-4408	Sequence 4408, Ap
43	6	2.8	257	9	US-09-764-884-23	Sequence 23, Appl
44	6	2.8	277	10	US-09-864-761-46564	Sequence 46564, A
45	6	2.8	281	9	US-09-764-884-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-10-024-955-7
; Sequence 7, Application US/10024955
; Patent No. US20020168373A1
; GENERAL INFORMATION:
; APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
; TITLE OF INVENTION: Allergenic Proteins and Peptides From House Dust Mite and Uses Therefor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/024,955
; FILING DATE: 19-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,336A
; FILING DATE: 10-JUN-1996
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMT-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-557-309B-40

Query Match 2.8%; Score 6; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 157 FEVROF 162
|||||
Db 46 FEVROF 51

Search completed: February 21, 2003, 10:16:39
Job time : 18 secs

STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/028,463
FILING DATE: 09-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/694,747
FILING DATE: 02-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7005-023-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-9741
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Rhodococcus rhodochrous
STRAIN: J-1 (FERM BP-1478)
US-08-028-463-4

Query Match 2.8%; Score 6; DB 1; Length 226;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NVVNH1 169
DB 211 NVVNH1 216

RESULT 49
US-08-461-836-4
Sequence 4, Application US/08461836
Patent No. 5753472
GENERAL INFORMATION:
APPLICANT: BEPPU, TERUHIKO
APPLICANT: YAMADA, HIDEAKI
APPLICANT: NAGASAWA, TORU
APPLICANT: HORINOCHI, SUCHAYU
APPLICANT: NISHIYAMA, MAKATO
TITLE OF INVENTION: DNA FRAGMENT ENCODING A POLYPEPTIDE
TITLE OF INVENTION: HAVING NITRILE HYDROLASE ACTIVITY, A TRANSFORMANT CONTAINING THE
TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
TITLE OF INVENTION: TRANSFORMANT
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/461,836
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/028,463
FILING DATE: 09-MAR-1993
APPLICATION NUMBER: US 07/694,747
FILING DATE: 02-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7005-023-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-9741
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Rhodococcus rhodochrous
STRAIN: J-1 (FERM BP-1478)
US-08-461-836-4

Query Match 2.8%; Score 6; DB 1; Length 226;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NVVNH1 169
DB 211 NVVNH1 216

RESULT 50
US-08-557-309B-40
Sequence 40, Application US/08557309B
Patent No. 5916572
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 683-6031
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 233 amino acids

```
US-09-446-959-10
; Sequence 10, Application US/09446959
; Patent No. 6426070
; GENERAL INFORMATION:
; APPLICANT: ROSENBERG, HELENE F.
; APPLICANT: DOMACHOWSKIE, JOSEPH B.
; TITLE OF INVENTION: METHODS FOR INACTIVATING ENVELOPED RNA VIRUS PARTICLES
; FILE REFERENCE: AND COMPOSITIONS FOR USE THEREWITH
; FILE REFERENCE: 11613.8USMO
; CURRENT APPLICATION NUMBER: US/09/446,959
; CURRENT FILING DATE: 2000-05-09
; PRIOR APPLICATION NUMBER: PCT/US98/13852
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/052,986
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-446-959-10

Query Match      2.8%; Score 6; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 FANVYN 167
DB 75 FANVYN 80

RESULT 45
US-09-230-637-36
; Sequence 36, Application US/09230637
; Patent No. 6264958
; GENERAL INFORMATION:
; APPLICANT: Hayward, Gary
; APPLICANT: Nicholas, John
; APPLICANT: Hardwick, J. Marie
; APPLICANT: Reitz, Marvin
; TITLE OF INVENTION: NO. 6264958el Genes of Kaposi's Sarcoma
; FILE REFERENCE: 1107.78372
; CURRENT APPLICATION NUMBER: US/09/230,637
; CURRENT FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 60/022,591
; PRIOR FILING DATE: 1996-07-25
; PRIOR APPLICATION NUMBER: PCT US 97/12931
; PRIOR FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-230-637-36

Query Match      2.8%; Score 6; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 GVLSDV 185
DB 164 GVLSDV 169

RESULT 46
US-09-134-001C-4319
; Sequence 4319, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

US-09-446-959-10
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4319
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4319

Query Match      2.8%; Score 6; DB 4; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 EYDLAY 120
DB 57 EYDLAY 62

RESULT 47
US-08-861-774E-44
; Sequence 44, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; FILE REFERENCE: BIOACTIVE MOLECULES
; CURRENT APPLICATION NUMBER: 9993-006
; CURRENT FILING DATE: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Parmelia sulcata
US-08-861-774E-44

Query Match      2.8%; Score 6; DB 4; Length 214;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATIAA 39
DB 157 DDATIAA 162

RESULT 48
US-08-028-463-4
; Sequence 4, Application US/08028463
; Patent No. 5731176
; GENERAL INFORMATION:
; APPLICANT: BEPPU, TERUHIKO
; APPLICANT: YAMADA, HIDEAKI
; APPLICANT: NAGASAWA, TORU
; APPLICANT: HORINOUCHI, SUDHAYU
; APPLICANT: NISHIYAMA, MAKATO
; TITLE OF INVENTION: DNA FRAGMENT ENCODING A POLYPEPTIDE
; TITLE OF INVENTION: HAVING NITRILE HYDROLASE ACTIVITY, A TRANSFORMANT CONTAININ
; TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
```

Db 1 MKFLI 6

RESULT 41

US-09-446-959-7
; Sequence 7, Application US/09446959
; Patent No. 6426070
; GENERAL INFORMATION:
; APPLICANT: ROSENBERG, HELENE F.
; APPLICANT: DOMACHOWSKI, JOSEPH B.
; TITLE OF INVENTION: METHODS FOR INACTIVATING ENVELOPED RNA VIRUS PARTICLES
; FILE REFERENCE: 11613.8USMO
; CURRENT APPLICATION NUMBER: US/09/446,959
; PRIOR APPLICATION NUMBER: PCT/US98/13852
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/052,986
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 7
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-446-959-7

Query Match 2.8%; Score 6; DB 4; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 FANVN 167
Db 49 FANVN 54

RESULT 42

US-08-460-739-3
; Sequence 3, Application US/08460739
; Patent No. 5698416
; GENERAL INFORMATION:
; APPLICANT: Wolf, Marcia K.
; APPLICANT: Cassels, Frederick J.
; APPLICANT: Bell, Brian A.
; TITLE OF INVENTION: Improved Methods for Production of
; TITLE OF INVENTION: Antigens Under Control of Temperature-Regulated Promoters
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glenna Hendricks
; STREET: 9669 A Main Street
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,739
; FILING DATE: 20-APR-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna M.
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: WOLF2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 170 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: CPA1 protein
US-08-460-739-3

Query Match 2.8%; Score 6; DB 1; Length 170;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FVAVA 17
Db 16 FVAVA 21

RESULT 43

US-08-483-101-11
; Sequence 11, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Farber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 170 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-483-101-11

Query Match 2.8%; Score 6; DB 2; Length 170;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 FVAVA 17
Db 16 FVAVA 21

RESULT 44

LENGTH: 134 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..134
OTHER INFORMATION: /note="Human eosinophil derived
US-08-891-848-14
neurotoxin (EDN)"
Query Match 2.8%; Score 6; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 162 FANVNV 167
|||||
DB 48 FANVNV 53
RESULT 38
US-08-875-811-9
Sequence 9, Application US/0875811
Patent No. 6045793
GENERAL INFORMATION:
APPLICANT: Rybak, Susanna M.
APPLICANT: Newton, Diane L.
APPLICANT: Boque, Luis
APPLICANT: Wlodawer, Alexander
TITLE OF INVENTION: Recombinant Ribonuclease Proteins
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,811
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/02588
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,800
FILING DATE: 21-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Faris, Susan K.
REGISTRATION NUMBER: 41,739
REFERENCE/DOCKET NUMBER: 015280-244100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 134 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..134
OTHER INFORMATION: /note="Human eosinophil-derived
OTHER INFORMATION: neurotoxin (EDN)"

US-08-875-811-9
Query Match 2.8%; Score 6; DB 3; Length 134;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 162 FANVNV 167
|||||
DB 48 FANVNV 53
RESULT 39
US-09-446-959-5
Sequence 5, Application US/09446959
Patent No. 6426070
GENERAL INFORMATION:
APPLICANT: ROSENBERG, HELENE F.
APPLICANT: DOMACHOWSKI, JOSEPH B.
TITLE OF INVENTION: METHODS FOR INACTIVATING ENVELOPED RNA VIRUS PARTICLES
FILE REFERENCE: 11613, 8USWO
CURRENT APPLICATION NUMBER: US/09/446,959
CURRENT FILING DATE: 2000-05-09
PRIOR APPLICATION NUMBER: PCT/US98/13852
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/052,986
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 134
TYPE: PRT
ORGANISM: Homo sapiens
US-09-446-959-5
Query Match 2.8%; Score 6; DB 4; Length 134;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 162 FANVNV 167
|||||
DB 48 FANVNV 53
RESULT 40
US-09-383-586-35
Sequence 35, Application US/09383586
Patent No. 6242419
GENERAL INFORMATION:
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Abernethy, Nevin
APPLICANT: Omrusl, Rene
APPLICANT: Kumble, Anand
TITLE OF INVENTION: Compounds isolated from stromal cells
FILE REFERENCE: 11000.1037c1
CURRENT APPLICATION NUMBER: US/09/383,586
CURRENT FILING DATE: 1999-08-26
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 35
LENGTH: 147
TYPE: PRT
ORGANISM: Mouse
US-09-383-586-35
Query Match 2.8%; Score 6; DB 4; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 MKFLLI 7

APPLICATION NUMBER: US 07/779,195
FILING DATE: 22-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/510,696
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 015280-110310US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..133
OTHER INFORMATION: /note="Human eosinophil cationic
OTHER INFORMATION: protein (ECP)"
US-08-891-848-15

Query Match 2.8%; Score 6; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 162 FANYVN 167
|||||
Db 48 FANYVN 53

RESULT 36
US-08-875-811-10
Sequence 10, Application US/08875811
Patent No. 6045793
GENERAL INFORMATION:
APPLICANT: Rybak, Susanna M.
APPLICANT: Newton, Dianne L.
APPLICANT: Bogue, Lluís
APPLICANT: Mlodawer, Alexander
TITLE OF INVENTION: Recombinant Ribonuclease Proteins
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,811
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/02588
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA: US 60/011,800
FILING DATE: 21-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Faris, Susan K.
REGISTRATION NUMBER: 41,739
REFERENCE/DOCKET NUMBER: 015280-244100US
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..133
OTHER INFORMATION: /note="Human eosinophil cationic
OTHER INFORMATION: protein (ECP)"
US-08-875-811-10

Query Match 2.8%; Score 6; DB 3; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 162 FANYVN 167
|||||
Db 48 FANYVN 53

RESULT 37
US-08-891-848-14
Sequence 14, Application US/08891848
Patent No. 5955073
GENERAL INFORMATION:
APPLICANT: Rybak, Susanna M.
APPLICANT: Youle, Richard J.
APPLICANT: Newton, Dianne L.
APPLICANT: Nicholls, Peter J.
TITLE OF INVENTION: Selective RNase Cytotoxic Reagents
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/891,848
FILING DATE: No. 5955073 yet assigned
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,462
FILING DATE: 22-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/014,082
FILING DATE: 04-FEB-1993
PRIOR APPLICATION DATA: US 07/779,195
FILING DATE: 22-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/510,696
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 015280-110310US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:

```
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 128
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-150

Query Match      2.8%; Score 6; DB 4; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 GGLSTL 175
Db 23 GGLSTL 28

RESULT 33
US-08-944-449-2
; Sequence 2, Application US/08944449
; Patent No. 5985613
; GENERAL INFORMATION:
; APPLICANT: KURTH, REINHARD
; APPLICANT: BAUER, MICHAEL
; APPLICANT: METZNER, KARIN
; APPLICANT: WERNER, ALBRECHT
; TITLE OF INVENTION: Use of an "immunodeficiency-virus suppressing
; TITLE OF INVENTION: lymphokine (ISL)" to inhibit the replication of
; TITLE OF INVENTION: viruses, particularly of retroviruses
; FILE REFERENCE: 8341-7065
; CURRENT APPLICATION NUMBER: US/08/944,449
; CURRENT FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: EP 95113013.2
; EARLIER FILING DATE: 1995-08-18
; EARLIER APPLICATION NUMBER: DE 195 13 152.5
; EARLIER FILING DATE: 1995-04-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 130
; TYPE: PRF
; ORGANISM: african green monkey
US-08-944-449-2

Query Match      2.8%; Score 6; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EOSETI 46
Db 71 EOSETI 76

RESULT 34
US-09-353-362-2
; Sequence 2, Application US/09353362
; Patent No. 6383739
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Use of an "immunodeficiency-virus suppressing
; TITLE OF INVENTION: lymphokine (ISL)" to inhibit the replication of viruses,
; TITLE OF INVENTION: in particular of retroviruses
; NUMBER OF SEQUENCES: 8
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/353,362
; FILING DATE: 15-JUL-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 13 152.5
; FILING DATE: 07-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95113013.2
; FILING DATE: 18-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KLESNER, Sharon N.
; REGISTRATION NUMBER: 36,335
; REFERENCE/DOCKET NUMBER: P8341-9012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-5000
; TELEFAX: (202) 638-4810
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 130 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-353-362-2

Query Match      2.8%; Score 6; DB 4; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EOSETI 46
Db 71 EOSETI 76

RESULT 35
US-08-891-848-15
; Sequence 15, Application US/08891848
; Patent No. 5955073
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Toule, Richard J.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Nicholls, Peter J.
; TITLE OF INVENTION: Selective RNase Cytotoxic Reagents
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,848
; FILING DATE: NO. 5955073 yet assigned
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,462
; FILING DATE: 22-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/014,082
; FILING DATE: 04-FEB-1993
; PRIOR APPLICATION DATA:
```

1 TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
2 NUMBER OF SEQUENCES: 503
3 CORRESPONDENCE ADDRESS:
4 ADDRESSEE: Knobbe, Martens, Olson & Bear
5 STREET: 501 West Broadway
6 CITY: San Diego
7 STATE: California
8 COUNTRY: USA
9 ZIP: 92101-3505
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy Disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: Win95
14 SOFTWARE: Word
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/905,223
17 FILING DATE:
18 CLASSIFICATION: 536
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Israelson, Ned A.
21 REGISTRATION NUMBER: 29,655
22 REFERENCE/DOCKET NUMBER:
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (619) 235-8550
25 TELEFAX: (619) 235-0176
26 INFORMATION FOR SEQ ID NO: 335:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 67 amino acids
29 TYPE: AMINO ACID
30 TOPOLOGY: LINEAR
31 MOLECULE TYPE: PROTEIN
32 ORIGINAL SOURCE:
33 ORGANISM: Homo Sapiens
34 TISSUE TYPE: Brain
35 FEATURE:
36 NAME/KEY: sig-peptide
37 LOCATION: -43...-1
38 IDENTIFICATION METHOD: Von Heijne matrix
39 OTHER INFORMATION: score 5.2
40 US-08-905-223-335

Query Match 2.8%; Score 6; DB 4; Length 67;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 VLSDVL 186
Db 37 VLSDVL 42

RESULT 29
US-09-383-586-34
Sequence 34, Application US/09383586
Patent No. 6242419
GENERAL INFORMATION:
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Abernethy, Nevin
APPLICANT: Onrust, Rene
APPLICANT: Kumble, Anand
APPLICANT: Murison, Greg
TITLE OF INVENTION: Compounds isolated from stromal cells
TITLE OF INVENTION: and methods for their use
FILE REFERENCE: 11000.1037c1
CURRENT APPLICATION NUMBER: US/09/383,586
CURRENT FILING DATE: 1999-08-26
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 102
TYPE: PRT
ORGANISM: Mouse

US-09-383-586-34
Query Match 2.8%; Score 6; DB 4; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MKFLLI 7
Db 1 MKFLLI 6

RESULT 30
US-09-383-586-34
Patent No. 5164490
APPLICANT: SANTU, DANIEL V.;EDMAN, JEFFREY;EDMAN, URSULA
TITLE OF INVENTION: PNEUMOCYSTIS CARINI DIIHYDROLYATE
REDUCTASE GENE AND METHODS FOR ITS USE
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/437,511
FILING DATE: 15-NOV-1989
SEQ ID NO:6:
LENGTH: 115
5164490-6

Query Match 2.8%; Score 6; DB 6; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 GVLSDV 185
Db 93 GVLSDV 98

RESULT 31
US-09-562-737-125
Sequence 125, Application US/09562737
Patent No. 6428967
GENERAL INFORMATION:
APPLICANT: Herz, Joachim
APPLICANT: Gotthardt, Michael
TITLE OF INVENTION: LDL Receptor Signaling Pathways
FILE REFERENCE: UTSW0708
CURRENT APPLICATION NUMBER: US/09/562,737
CURRENT FILING DATE: 2000-05-01
NUMBER OF SEQ ID NOS: 132
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 125
LENGTH: 116
TYPE: PRT
ORGANISM: mouse
US-09-562-737-125

Query Match 2.8%; Score 6; DB 4; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 VALSLE 145
Db 106 VALSLE 111

RESULT 32
US-09-199-637A-150
Sequence 150, Application US/09199637A
Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wah
APPLICANT: Cao, Hui

RESULT 25
PCT-US95-09281-3
; Sequence 3, Application PC/TUS9509281
; GENERAL INFORMATION:
; APPLICANT: Cytotherapeutics, Inc.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF GROWTH CONTROL
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., FISH & NEAVE
; STREET: 1251 Ave. of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09281
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,773
; FILING DATE: 20-JULY-1994
; APPLICATION NUMBER: US 08/432,698
; FILING DATE: 09-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: CTI-22 CIP PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US95-09281-3

Query Match 2.8%; Score 6; DB 5; Length 19;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 13 VAVSAD 18

RESULT 26
PCT-US95-09282-3
; Sequence 3, Application PC/TUS9509282
; GENERAL INFORMATION:
; APPLICANT: Cytotherapeutics, Inc.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR A
; TITLE OF INVENTION: BIOARTIFICIAL EXTRACELLULAR MATRIX
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., FISH & NEAVE
; STREET: 1251 Ave. of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09282
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,646
; FILING DATE: 20-JULY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: CTI-26 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US95-09282-3

Query Match 2.8%; Score 6; DB 5; Length 19;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 13 VAVSAD 18

RESULT 27
US-09-562-737-124
; Sequence 124, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
; LENGTH: 57
; TYPE: PPT
; ORGANISM: mouse
; US-09-562-737-124

Query Match 2.8%; Score 6; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 VALSLE 145
Db 47 VALSLE 52

RESULT 28
US-08-905-223-335
; Sequence 335, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno

TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-279-773-3

Query Match 2.8%; Score 6; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 13 VAVSAD 18

RESULT 23
US-08-747-137-56
; Sequence 56, Application US/08747137
; Patent No. 5945033
; GENERAL INFORMATION:
; APPLICANT: YEN, Richard C.K.
; TITLE OF INVENTION: NON-CROSSLINKED PROTEIN PARTICLES FOR
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC USE
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,137
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,546
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/069,831
; FILING DATE: 01-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/959,560
; FILING DATE: 13-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/641,720
; FILING DATE: 15-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 016197-000840US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
US-08-747-137-56

Query Match 2.8%; Score 6; DB 2; Length 19;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 13 VAVSAD 18

RESULT 24
US-08-447-997-3
; Sequence 3, Application US/08447997
; Patent No. 6392118
; GENERAL INFORMATION:
; APPLICANT: Schinstine, Malcolm
; APPLICANT: Scholchet, Molly S.
; APPLICANT: Gentile, Frank T.
; APPLICANT: Hamman, Joseph P.
; APPLICANT: Holland, Laura M.
; APPLICANT: Doherty, Edward J.
; APPLICANT: Winn, Shelley R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF GROWTH CONTROL
; TITLE OF INVENTION: FOR CELLS ENCAPSULATED IN BIOARTIFICIAL ORGANS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., FISH & NEAVE
; STREET: 1251 Ave. of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,997
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,698
; FILING DATE: 09-MAY-1995
; APPLICATION NUMBER: US 08/279,773
; FILING DATE: 20-JULY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: CTI-22 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-447-997-3

Query Match 2.8%; Score 6; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 13 VAVSAD 18

STATE: New York
COUNTRY: USA
ZIP: 10020-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,698
FILING DATE: 09-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,773
FILING DATE: 20-JULY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: CTT-22 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-432-698-3

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSAD 18
Db 13 VAVSAD 18

RESULT 21
US-08-447-810-3
Sequence 3, Application US/08447810
Patent No. 5858747
GENERAL INFORMATION:
APPLICANT: Schinstine, Malcolm
APPLICANT: Shoichet, Molly S.
APPLICANT: Gentile, Frank T.
APPLICANT: Hamman, Joseph P.
APPLICANT: Holland, Laura M.
APPLICANT: Cain, Brian M.
APPLICANT: Doherty, Edward J.
APPLICANT: Winn, Shelley R.
APPLICANT: Aebischer, Patrick
TITLE OF INVENTION: METHODS AND COMPOSITIONS OF GROWTH CONTROL
TITLE OF INVENTION: FOR CELLS ENCAPSULATED IN BIOARTIFICIAL ORGANS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: James F. Haley, Jr., FISH & NEAVE
STREET: 1251 Ave. of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,810

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/432,698
FILING DATE: 09-MAY-1995
APPLICATION NUMBER: US 08/279,773
FILING DATE: 20-JULY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: CTT-22 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-447-810-3

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSAD 18
Db 13 VAVSAD 18

RESULT 22
US-08-279-773-3
Sequence 3, Application US/08279773
Patent No. 5935849
GENERAL INFORMATION:
APPLICANT: Schinstine, Malcolm
APPLICANT: Shoichet, Molly S.
APPLICANT: Gentile, Frank T.
APPLICANT: Hamman, Joseph P.
APPLICANT: Holland, Laura M.
APPLICANT: Cain, Brian M.
APPLICANT: Doherty, Edward J.
APPLICANT: Winn, Shelley R.
APPLICANT: Aebischer, Patrick
TITLE OF INVENTION: METHODS AND COMPOSITIONS OF GROWTH CONTROL
TITLE OF INVENTION: FOR CELLS ENCAPSULATED IN BIOARTIFICIAL ORGANS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: James F. Haley, Jr., FISH & NEAVE
STREET: 1251 Ave. of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/279,773
FILING DATE: 20-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: CTT-22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000

QY 13 VAVSAD 18
|11111|
Db 13 VAVSAD 18

RESULT 18
US-08-280-646-3

; Sequence 3, Application US/08280646
; Patent No. 5834029

GENERAL INFORMATION:

APPLICANT: Bellamkonda, Ravi

APPLICANT: Raniert, John P

APPLICANT: Aebischer, Patrick

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR A

TITLE OF INVENTION: BIOARTIFICIAL EXTRACELLULAR MATRIX

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: James F. Haley, Jr., FISH & NEAVE

STREET: 1251 Ave. of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10020-1104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/280,646

FILING DATE: 20-JUL-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: CTI-26

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 19 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-280-646-3

Query Match 2.8%; Score 6; DB 2; Length 19;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
|11111|
Db 13 VAVSAD 18

RESULT 19
US-08-445-193-3

; Sequence 3, Application US/08445193
; Patent No. 5840576

GENERAL INFORMATION:

APPLICANT: Schinstine, Malcolm

APPLICANT: Shoichet, Molly S.

APPLICANT: Gentile, Frank T.

APPLICANT: Hamman, Joseph P.

APPLICANT: Holland, Laura M.

APPLICANT: Cain, Brian M.

APPLICANT: Doherty, Edward J.

APPLICANT: Winn, Shelley R.

APPLICANT: Aebischer, Patrick
TITLE OF INVENTION: METHODS AND COMPOSITIONS OF GROWTH CONTROL
TITLE OF INVENTION: FOR CELLS ENCAPSULATED IN BIOARTIFICIAL ORGANS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

ADDRESSEE: James F. Haley, Jr., FISH & NEAVE

STREET: 1251 Ave. of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10020-1104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/445,193

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/432,698

FILING DATE: 09-MAY-1995

APPLICATION NUMBER: US 08/279,773

FILING DATE: 20-JULY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: CTI-22 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 19 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-445-193-3

Query Match 2.8%; Score 6; DB 2; Length 19;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
|11111|
Db 13 VAVSAD 18

RESULT 20
US-08-432-698-3

; Sequence 3, Application US/08432698
; Patent No. 5843431

GENERAL INFORMATION:

APPLICANT: Schinstine, Malcolm

APPLICANT: Shoichet, Molly S.

APPLICANT: Gentile, Frank T.

APPLICANT: Hamman, Joseph P.

APPLICANT: Holland, Laura M.

APPLICANT: Cain, Brian M.

APPLICANT: Doherty, Edward J.

APPLICANT: Winn, Shelley R.

APPLICANT: Aebischer, Patrick

TITLE OF INVENTION: METHODS AND COMPOSITIONS OF GROWTH CONTROL

TITLE OF INVENTION: FOR CELLS ENCAPSULATED IN BIOARTIFICIAL ORGANS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: James F. Haley, Jr., FISH & NEAVE

STREET: 1251 Ave. of the Americas

CITY: New York

COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/895,252
FILING DATE: 19920608
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/646,291
FILING DATE: 25-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kowalchuk, Alan W.
REGISTRATION NUMBER: 31,535
REFERENCE/DOCKET NUMBER: 600.212-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Synthetically derived
US-07-895-252-10

Query Match 2.8%; Score 6; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
DB 12 VAVSAD 17

RESULT 16
US-08-072-283B-10
Sequence 10, Application US/08072283B
Patent No. 5703205
GENERAL INFORMATION:
APPLICANT: Skuditz, Amy P.N.
TITLE OF INVENTION: FURCHL, Leo T.
TITLE OF INVENTION: LAMININ A CHAIN POLYPEPTIDES
TITLE OF INVENTION: FROM THE AMINO TERMINAL GLOBULAR DOMAIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 3100 No. 5703205west Center, 90 South Seventh St
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,283B
FILING DATE: 07-JUN-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/895,252
FILING DATE: 08-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:

REFERENCE/DOCKET NUMBER: 600.212-US-02
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 5703205e
US-08-072-283B-10

Query Match 2.8%; Score 6; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
DB 12 VAVSAD 17

RESULT 17
US-07-998-820-2
Sequence 2, Application US/07998820
Patent No. 5738838
GENERAL INFORMATION:
APPLICANT: Zamora, Paul O.
TITLE OF INVENTION: IKVAV Peptide
TITLE OF INVENTION: Radiopharmaceutical Applications
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhomed Incorporated
STREET: 4261 Balloon Park
CITY: Albuquerque
STATE: NM
COUNTRY: U.S.A.
ZIP: 87109-5802
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
MEDIUM TYPE: Storage
COMPUTER: IBM PC/XT/AT, IBM PS/2 or
COMPUTER: compatibles
OPERATING SYSTEM: PC-DOS or MS-DOS
SOFTWARE: WordPerfect 6.0a for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/998,820
FILING DATE: 30-DEC-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/840,077
FILING DATE: 20-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: David P. Hegge
REGISTRATION NUMBER: 36,827
TELECOMMUNICATION INFORMATION:
TELEPHONE: (505) 344-7200
TELEFAX: (505) 344-9460
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-998-820-2

Query Match 2.8%; Score 6; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 494-150-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-405-200-2

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSAD 18
Db 6 VAVSAD 11

RESULT 13
US-08-405-200-15
Sequence 15, Application US/08405200
Patent No. 5696229
GENERAL INFORMATION:
APPLICANT: LAURIE, Gordon W
APPLICANT: MATTER, Michelle L
APPLICANT: CHEN, Ianlin
TITLE OF INVENTION: POLYPEPTIDE WITH LAMININ CELL ADHESION
TITLE OF INVENTION: AND MORPHOGENESIS ACTIVITY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,200
FILING DATE: 16-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 494-150-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-405-200-15

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSAD 18
Db 6 VAVSAD 11

RESULT 14
PCT-US91-08497-54
Sequence 54, Application PC/TUS9108497
GENERAL INFORMATION:
APPLICANT: Fleming, Patrick J.
APPLICANT: Kent, Ute M.
TITLE OF INVENTION: Peptide which Regulates Weight Gain in Mammals
TITLE OF INVENTION: Mammals
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08497
FILING DATE: 19911121
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/616,910
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, Norman F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-069-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US91-08497-54

Query Match 2.8%; Score 6; DB 5; Length 12;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 YDLAYK 121
Db 2 YDLAYK 7

RESULT 15
US-07-895-252-10
Sequence 10, Application US/07895252
Patent No. 5276136
GENERAL INFORMATION:
APPLICANT: Skubitz, Amy P.N.
APPLICANT: Furcht, Leo T.
TITLE OF INVENTION: LAMININ A CHAIN POLYPEPTIDES FROM THE AMINO TERMINAL GLOBULAR DOMAIN
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5276136 West Center
CITY: Minneapolis
STATE: MN

TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-029-333-31

Query Match 2.8%; Score 6; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
DB 4 VAVSAD 9

RESULT 10
US-08-391-820-31
Sequence 31, Application US/08391820
Patent No. 6469138
GENERAL INFORMATION:
APPLICANT: Frazier, William A.
Kosfeld, Minh D.
TITLE OF INVENTION: Thrombospondin Receptor Binding Peptides
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SG
STREET: 800 N. Lindbergh Blvd.
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63167
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/391,820
FILING DATE: 21-Feb-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/029,333
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meyer, Scott J.
REGISTRATION NUMBER: 25,275
REFERENCE/DOCKET NUMBER: 07-24(982)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)694-3117
TELEFAX: (314)694-5435
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-08-391-820-31
Query Match 2.8%; Score 6; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
DB 4 VAVSAD 9

RESULT 11
US-07-616-910-54
Sequence 54, Application US/07616910
Patent No. 5223464
GENERAL INFORMATION:
APPLICANT: Fleming, Patrick J.

APPLICANT: Kent, Ute M.
TITLE OF INVENTION: Peptide which Regulates Weight Gain in
Mammals
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/616,910
FILING DATE: 19901121
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: OBION, NO. 5223464man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-069-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-616-910-54
Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YDLAYK 121
DB 2 YDLAYK 7

RESULT 12
US-08-405-200-2
Sequence 2, Application US/08405200
Patent No. 5696229
GENERAL INFORMATION:
APPLICANT: LAURIE, Gordon W
MATTER, Michelle L
APPLICANT: CHEN, Lanlin
TITLE OF INVENTION: POLYPEPTIDE WITH LAMININ CELL ADHESION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MATER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,200
FILING DATE: 16-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

DB 730 KRISSEYNNYEMAKNFSLSIDITAEFEVGRKEIKLGNINIDS--IKSEYEISER 787
QY 169 IGGSLIDPIFGVLSVDVLAIFODTVR---KEMTKYVL 202
DB 788 YDGLKATE-----SEIVDA--KDKIESAIKEMDKII 816

RESULT 9

Q9PEP2 PRELIMINARY; PRT; 547 AA.
AC Q9PEP2: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE 60 kDa chaperonin (protein CPN60) (GroEL protein).
GN XFE015.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facchini A.P., Ferreira A.J.S., Ferreira V.C.A., Fierro J.A.,
RA Fraga J.S., Franco S.C., Franco M.C., Frohme M., Furian L.R.,
RA Gartner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nham J.A., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peloto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawaaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Melandis J., Setubal J.C.,
RT "The genome sequence of the plant pathogen Xylella fastidiosa";
RL Nature 406:151-159(2000).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AE003907; AAF83425.1; -.
DR HSSP: P06139; IGRU.
DR InterPro: IPR001843; Chaperin_Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone; Complete proteome.
SQ SEQUENCE 547 AA; 57757 MW; 30BE7P937CA7A9D8 CRC64;

Query Match 8.5%; Score 90.5; DB 16; Length 547;
Best Local Similarity 22.8%; Pred. No. 34;
Matches 54; Conservative 43; Mismatches 87; Indels 53; Gaps 12;

QY 13 VAVSADPIHDKITTEINKAIDDA-IAIPOSETIDPMKVPDADKFERHVGIVDFKGL 71
DB 107 VAAGMNPM-----DLKRIDKAIVAAVTELKISK---PTSDDKAIAQAVAIISANSDE 156
QY 72 AMRNIARGLKOKKROGDANVKGEGIVAKHLLIGVHDDIVSMEDLAY----- 120
DB 157 SIGNIIIAEMKKVKGEGVITI--EEGTLTLENL----DVEGQDFRGYSSPYFINNOQS 210
QY 121 ---KIGDLHPTTH--VTSIDQFVVALSLEISDEGNTMTSFEVROFA---NVNHIIGL 172
DB 211 QIVELDNPTYLHDKKISSVRDLITVLDAAVAKSKPLIYAEEVEGEALATLVVNNIRGI 270
QY 173 ----SIIDPIFG-----VLSVDL-----FAIFOD---TVRKMTKYLAFAFKRELEK 212
DB 271 IKVCAVKAQFGDRRKAMLEDAVNLGTGYISEVGLSLEKATTSHLGRKKAVYSK 327

RESULT 10

Q9L7P5 PRELIMINARY; PRT; 548 AA.
AC Q9L7P5: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE 60 kDa chaperonin (protein Cpn60) (GroEL protein).
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RA Wong H.-C., Lu K.-H.;
RT Submitted (JAN-2000) to the EMBL/Genbank/DBD databases.
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AF221845; AAF27528.1; -.
DR HSSP: P06139; IGRU.
DR InterPro: IPR001843; Chaperin_Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR Pfam: PF00118; cpn60_TCP1; 1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 548 AA; 57629 MW; E39223E289D6F74C CRC64;

Query Match 8.5%; Score 90.5; DB 2; Length 548;
Best Local Similarity 20.9%; Pred. No. 34;
Matches 49; Conservative 40; Mismatches 82; Indels 63; Gaps 9;

QY 13 VAVSADPIHDKITTEINKAIDDAI-AIEOSEITPMKVPDADKFERHVGIVDFKGL 71
DB 107 VAAGMNPM-----DLKRIDKAIVAAVQ---LKELSYECNDTIAIAQVIGISANSDA 156
QY 72 AMRNIARGLKOKKROGDANVKGEGIVAKHLLIGVHDDIVSMEDLAY----- 120
DB 157 SVGNIIIAEMERVGDRGVITV--EEG---QALODELDVVEGQDFRGYLSPYFINNOQA 210
QY 121 -----KIGDLHPTTHVTSIDQFVVALSLEISDEGNTMTSFEVROF 162
DB 211 GSVLENPPILLVDKKSISNRELLPLEVAAKASRPLLIITAEVGEALATL----- 262
QY 163 ANYVNIIGL----SIIDPIFGVLSVDVLAIFODTVRKEMTKYLAFAFKRELEK 212
DB 263 --VVNNMIRGIYKAAVKAQFGF---DRRAMQDIAIILGTGYISEIGIELEK 311

RESULT 11
Q18126 PRELIMINARY; PRT; 681 AA.
ID Q18126

```

AC Q18126;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hypothetical 79.5 kDa protein.
CN C2A43.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; Pubmed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A.;
RT "The sequence of C. elegans cosmid C2A43 ";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U04024; AAA81461.1; -.
KW Hypothetical protein.
SQ SEQUENCE 681 AA; 79533 MW; 2DA37C9046C26A20 CRC64;

```

Query March 8.5%; Score 90.5; DB 5; Length 681;
 Best Local Similarity 25.1%; Pred. NO. 46;
 Matches 51; Conservative 41; Mismatches 62; Indels 49; Gaps 11;

```

QY 24 KITTEINKAID--ATAIE-----OSETI-----DPKVPDHDKFERHVGIVDFKGL 71
DB 133 KMTENQOKAKDFFNSRISVLESMSIRAOSEIRIVANENEMR-----HSFPAKLAEISGQLEL 187
QY 72 AMRNIEARGLKQMRQGDANVKGEGIVKAHLIGVHDDIVSMEDLAYKLGDLHPPTHV 131
DB 188 AMKOITSEKAKOKER-----FOKVNALALEHHL--ELGNSKIDKLM 228
QY 132 ISDIODFVALSLEISDEGNIT-MTSEFVRQFANVNVNIGLS--ILDPIFGVLSDVLT- 188
DB 229 NSEIQ-----ARKLHEGGLAKMTDIEDR-----VNNIVGGMKNSIDEMKNGKNVHMPA 278
QY 189 IFODTVREKMTKVLAPAFKRELE 211
DB 279 LDTDLALREMEALIAADKKLSME 301

```

```

RESULT 12
O96275 PRELIMINARY; PRT; 1558 AA.
AC 096275;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE RESA-H3 antigen.
GN PFE0915W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; Pubmed=9804551;
RA Gardner M.J., Tetstelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perle M.,

```

```

RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL; AE001424; AAC71972.1; -.
SQ SEQUENCE 1558 AA; 175658 MW; E2FC68E9036C81BC CRC64;

```

Query Match 8.4%; Score 90; DB 5; Length 1558;
 Best Local Similarity 23.3%; Pred. NO. 1.5e+02;
 Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;

```

QY 22 YDKITEINKAIDAIAIEOSETIIDPKVPDHDKFERHVGIVDFKGLAMRNIEARGL 81
DB 747 FNTVLDKVEETVEISGESLENNE---MDKAFSEIFDYNVGIOENLLTGFRSIEISIV 802
QY 82 KOMKROGDANVKGEGIVKAHLIGVHDDIVSMEDLAYKLGDLHPPTHVSDIQDFVVA 141
DB 803 IQSEKVDLN---ENVVSSIL-----DNIENKEGLLNKLENISSTEGVOETVTEHV-- 851
QY 142 LSLBISDEGNITMTSEFV---ROFANVNVNIGLS---ILDPIFGVLSDVLT-IFQ 191
DB 852 -----EQNV-YVDVDPAPAKDQFLGTLNAGGLKEMFNLDEVPKSESQVITVEELKD 903
QY 192 DTVRKEMTKVLAPAFKRELEKN 213
DB 904 EPVOKEVEKETVSIIE-EMEEN 924

```

```

RESULT 13
O900P0 PRELIMINARY; PRT; 1786 AA.
AC 0900P0;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Liver stage antigen-3 precursor.
GN LSA-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;
RT "LSA-3, a conserved pre-erythrocytic malaria antigen can induce
RT protection in chimpanzees.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007010; CAB65343.1; -.
DR InterPro; IPR001313; Pumilio/Puf.
KW Signal.
FT SIGNAL 64 1786
FT CHAIN 1 63 POTENTIAL.
SQ SEQUENCE 1786 AA; 200101 MW; 5DF536D7B5B1BD98 CRC64;

```

Query Match 8.4%; Score 90; DB 5; Length 1786;
 Best Local Similarity 23.3%; Pred. NO. 1.8e+02;
 Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;

```

QY 22 YDKITEINKAIDAIAIEOSETIIDPKVPDHDKFERHVGIVDFKGLAMRNIEARGL 81
DB 979 FNTVLDKVEETVEISGESLENNE---MDKAFSEIFDYNVGIOENLLTGFRSIEISIV 1034
QY 82 KOMKROGDANVKGEGIVKAHLIGVHDDIVSMEDLAYKLGDLHPPTHVSDIQDFVVA 141
DB 1035 IQSEKVDLN---ENVVSSIL-----DNIENKEGLLNKLENISSTEGVOETVTEHV-- 1083
QY 142 LSLBISDEGNITMTSEFV---ROFANVNVNIGLS---ILDPIFGVLSDVLT-IFQ 191
DB 1084 -----EQNV-YVDVDPAPAKDQFLGTLNAGGLKEMFNLDEVPKSESQVITVEELKD 1135
QY 192 DTVRKEMTKVLAPAFKRELEKN 213

```


Db 1136 EPVQKEVETVSIIE-EMEEN 1156

RESULT 14

09SVN7 PRELIMINARY: PRT: 2137 AA.
 ID 09SVN7
 AC 09SVN7
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
 DE Hypothetical 240.1 kDa protein.
 GN F1845.140 OR A14G13750.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
 OC NCBI_Taxid=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Bevan M., Weber N., Grueninger D., Schmidheini T., Bancroft I.,
 RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RA Weber N., Grueninger D., Schmidheini T., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN (4)
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL035528; CAB36839.1; -
 DR EMBL: AL161537; CAB78417.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 2137 AA: 240107 MW: 0A3AFCB275C2BC2 CRC64;

Query Match 8.3%; Score 88.5; DB 10; Length 2137;
 Best local similarity 21.8%; Pred. No. 3e+02;
 Matches 45; Conservative 34; Mismatches 82; Indels 45; Gaps 10;

QY 35 DATAIEOSTIDPMKVPDADK-----FERHVGIVDFKGLAM--RRIEAR-GLKQMKRQ 87
 DB 614 DSHFLELVQNAADNKNYPPEHPTLTFTLQKTGIIVLNNCGFMPENIRALCDVGQSTKK 673
 QY 88 GDANVKGEGGI-VKAMHLIG-----VHDDIVSMEDLAY-KLGDLPPT--HYISDIQDF 138
 DB 674 GSGGIYGGKIGCKFSYFRVSDAPEIHNSGHRFKFDISEGIGYILPTVPPHDISSLSM 733
 QY 139 VVALSEISDEG--NITMTSFEVROFANVYVNHIGLSIDLPF----- 179
 DB 734 LSGRALHLKAGWNTCTILPFRAIDSERTVNH-----EPMSDLHPSLFLFLRLQC 787
 QY 180 ----GVLSDVLTALFQDITVAKEMTKV 201
 DB 788 IYRNVLDLSDLVMRKEVVSKNITKV 813

RESULT 15

097PC9 PRELIMINARY: PRT: 436 AA.
 ID 097PC9
 AC 097PC9
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Phosphoglycerate dehydrogenase-related protein.
 GN SPI709.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacilliales;
 OC Streptococcaceae; Streptococcus.

OX NCBI_Taxid=1313;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-TIGR4;
 RX MEDLINE-21357209; PubMed-11463916;
 RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heideberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umamam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae";
 RT Science 293:498-506(2001).
 RL Science 293:498-506(2001).
 DR EMBL: AE007464; AAK75787.1; -
 DR TIGR: SP1709;
 DR InterPro: IPR005289; GTP-binding_dom.
 DR InterPro: IPR000765; GTP_OBG.
 DR InterPro: IPR002917; MMR_HSR1.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam: PF01926; MMR_HSR1; 1.
 DR PRINTS: PR00326; GTP_OBG.
 DR TIGRFAMS: TIGR00650; MG442; 2.
 DR TIGRFAMS: TIGR00231; small_GTP; 2.
 KW Complete proteome.
 SQ SEQUENCE 436 AA: 49082 MW: B013890D5285BBD CRC64;

Query Match 8.2%; Score 88; DB 16; Length 436;
 Best local similarity 21.8%; Pred. No. 41;
 Matches 46; Conservative 34; Mismatches 89; Indels 42; Gaps 8;

QY 30 NKAIDDAIAIEOSETIDPMKVPDADKFERHVGIVDFKGLAMRIEARGLKQMKRQD 89
 DB 21 NRIAGERISIVEDVEGTRDIYATGEMLRSMIDTGC---IDVDAPFMQIKHOAE 77
 QY 90 AN-----VKGEGYVK-----AHLLIGVHDDIV-----SMEYDI-----AYK 121
 DB 78 IAMEADVIVFVSGKEGITDADEYVARKLYKTKHPYILAVKNVDPENKNDIYDFYALG 137
 QY 122 LGDLPHTTHV-----ISDIDQFVVALSLEISDEGNITWTSFEVROFANV-----VNHIGL 172
 DB 138 LGEPRLPSSVHGIGTGVDIAIVENLPNEYEEENPDVIRKSLIGRPVNGSSLLMILGE 197
 QY 173 S--ILDPFGLSDVLTALFQDITVAKEMTKV 201
 DB 198 DRVIASPVAGTTRDAIDHTFTDQGEFTMI 228

RESULT 16

09ZVB0 PRELIMINARY: PRT: 372 AA.
 ID 09ZVB0
 AC 09ZVB0
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE F9K20.3.
 GN F9K20.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
 OC NCBI_Taxid=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,
 RA Kremenetskaia I., Luros J., Araujo R., Buchler E., Conway A.B.,
 RA Dewar K., Feng J., Kim C., Li Y., Shim P., Davis R.W., Ecker J.R.,
 RA Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F9K20 sequence";
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
 DE Hypothetical protein PF0807.
 GN PF0807.
 OS *Pyrococcus furiosus*.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC *Pyrococcus*.
 NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the *Pyrococcus furiosus* genome";
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE010197; AAL80931.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 562 AA; 64247 MW; 16EA3FE3946B80FA CRC64;

Query Match 8.1%; Score 87; DB 17; Length 562;
 Best Local Similarity 22.7%; Pred. No. 68;
 Matches 51; Conservative 36; Mismatches 70; Indels 68; Gaps 12;

QY 22 YDKITEINKAIDDAIAIEQSETIDPMKVPDHADKFERHVG-IVD-----FKG----- 69
 DB 63 YERLTSTILFLIE-----EONERLNFLPP--FLEFHRGIGIKIIGQRAFFKGMKSGV 114
 QY 70 -----ELAMRNIEARGLKQMKRGDANY-KGEEGIVKAKHLIGVHDDIVSMHEYDL 118
 DB 115 AMMKEGISESELALSNIESLEFTEDEKRLNDVKDKLERKVELSALEFLARRKVN 174
 QY 119 AVKGLDPTTHIVISDIODFVALSLSEISDEGNITMTSEFROFANVNHIGLSI---- 174
 DB 175 -----PLVLXVSDENFPIYE-----NVTFYGI-ARNFSEVILHIGNESKVKV 216
 QY 175 -----LDPIF-----GV-----LSDLVLAIFODTVRKEMTKVLAP 204
 DB 217 VEGESFLDYIFVSPGYIAVGAAGNVKSNVIRINVTIKIPRIIAP 261

RESULT 20

Q9AZS0 PRELIMINARY; PRT; 1640 AA.

AC 09AZS0;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 19, last annotation update)
 DE Tail protein.
 GN ORF52.
 OS bacteriophage bIL286.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 NCBI_TaxID=151536;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21113149; PubMed=11160885;
 RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
 RT "Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations";
 RL Nucleic Acids Res. 29:644-651(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF323669; NAK08339.1; -
 SQ SEQUENCE 1640 AA; 176321 MW; CEE59A7A47B8F18 CRC64;

Query Match 8.1%; Score 87; DB 9; Length 1640;
 Best Local Similarity 20.7%; Pred. No. 2.8e+02;
 Matches 47; Conservative 45; Mismatches 73; Indels 62; Gaps 12;

QY 16 SADPIHYDKITEINKAIDDAIAIEQSETIDPMKVPDHADKFERHVGIVDFKGLAMRN 75
 DB 1142 SMSKAOYDEIVKNAOKORDDTISAARKOQ---EVTDKAOK--THDKTVELANSKADKN 1195

QY 76 IEA-----RGLKMKRGDANYKGEIGIVK-----AHLLI----- 105
 DB 1196 VKAAAKGEGETVQYTKGRFSDNRNLNSPFDINGVNLFLHKMGNGIGHVSLKGFATGR 1255
 QY 106 GVHDDIVSMHEYDLAVKLGDLHPTTHIVIS-----DIODFVALSL---EISDEGNITMTS 156
 DB 1256 GLAODETALVGEGFELAH-HPSRGIFAVGQOGPEIRNLKAGTSLIPHSMSKE-FLSLT- 1312
 QY 157 FEVROFANVNHIGLSILDPITFGVSDVLTAIFODTVRKEMTKVLA 203
 DB 1313 -----ANLPAHADGV-----GFLSDALGVW--KSTYKDYVTSVIS 1345

RESULT 21

Q9CFST7 PRELIMINARY; PRT; 1640 AA.

AC 09CFST7;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE Prophage p13 protein 14.
 GN PF134 OR L1338.
 OS *Lactococcus lactis* (subsp. *lactis*) (*Streptococcus lactis*).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC *Streptococcaceae*; *Lactococcus*.
 NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RA MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403";
 RL Genome Res. 11:731-753(2001).
 DR EMBL: AE006370; AAK05486.1; -
 KW Complete proteome.
 SQ SEQUENCE 1640 AA; 176320 MW; CEE59A7A47B8F18 CRC64;

Query Match 8.1%; Score 87; DB 16; Length 1640;
 Best Local Similarity 20.7%; Pred. No. 2.8e+02;
 Matches 47; Conservative 45; Mismatches 73; Indels 62; Gaps 12;

QY 16 SADPIHYDKITEINKAIDDAIAIEQSETIDPMKVPDHADKFERHVGIVDFKGLAMRN 75
 DB 1142 SMSKAOYDEIVKNAOKORDDTISAARKOQ---EVTDKAOK--THDKTVELANSKADKN 1195
 QY 76 IEA-----RGLKMKRGDANYKGEIGIVK-----AHLLI----- 105
 DB 1196 VKAAAKGEGETVQYTKGRFSDNRNLNSPFDINGVNLFLHKMGNGIGHVSLKGFATGR 1255
 QY 106 GVHDDIVSMHEYDLAVKLGDLHPTTHIVIS-----DIODFVALSL---EISDEGNITMTS 156
 DB 1256 GLAODETALVGEGFELAH-HPSRGIFAVGQOGPEIRNLKAGTSLIPHSMSKE-FLSLT- 1312
 QY 157 FEVROFANVNHIGLSILDPITFGVSDVLTAIFODTVRKEMTKVLA 203
 DB 1313 -----ANLPAHADGV-----GFLSDALGVW--KSTYKDYVTSVIS 1345

RESULT 22

Q9KNR7 PRELIMINARY; PRT; 544 AA.

AC 09KNR7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE 60 kDa chaperonin (Protein CPN60) (GroEL protein).
 GN VC2664.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 NCBI_TaxID=666;
 RN [1]

```

RP SROUNCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406683; PubMed=10952301;
RA Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Esmolaeva M.D., Vamathevan J., Bass S., Qin H., Dracoi I., Sellers P.,
RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC - FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC - SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AE004332; AAF95805.1; -.
DR HSSP: P06139; 1GRL.
DR TIGR: VCB264; -.
DR InterPro: IPR001844; Chaperonin Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR Pfam: PF00118; Cpn60_TCP1.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONIN60_CPN60_1.
KW ATP-binding; Chaperone; Complete proteome.
SQ SEQUENCE 544 AA; 57153 MW; 927868AE794AE079 CRC64;

Query Match 8.1%; Score 86.5; DB 16; Length 544;
Best Local Similarity 24.4%; Pred. No. 72;
Matches 41; Conservative 21; Mismatches 69; Indels 37; Gaps 5;

QY 6 LIAAVAVAVASA-----DPIYDKTEELINKAIDAIAGSETIDPMKV 51
DB 372 LAGGAATVATKGAATVEEMKEKRDVEDALHATRAAVEGGVAGGVALIRAAKSLSLV- 430
QY 52 PDHAKFERHGVDFKGLAMRNIEARGLKOMKRGDA-----NKGEGGIYKAHLI 105
DB 431 ---GDNEQNVGI-----RALRAMEARPLRGIVYKNAGDEEVANNVRAGNGNYNAAT 482
QY 106 GVHDDIVSMEDLAKGLDHPHTTVISDIQDEVVALSLEISDEGNT 153
DB 483 GYVGDMIEIEM-----GIDPTKVTSAIDQAFASVAGLMTTAMIT 522

RESULT 23
QY1432 PRELIMINARY; PRT; 303 AA.
AC QY1432;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Probable transcriptional regulator.
GN PA1312.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN-ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Ralzer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen."

```

```

RL Nature 406:959-964(2000).
CC - SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL: AE004560; AAG04701.1; -.
DR InterPro: IPR000847; HTH_LYSR.
DR InterPro: IPR005119; LysR_subst.
DR Pfam: PF00126; HTH_1.1.
DR Pfam: PF03466; LysR_substrate_1.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
KW DNA-binding; Transcription regulation; Complete proteome.
SQ SEQUENCE 303 AA; 33062 MW; FE8C201582A0BF94 CRC64;

Query Match 8.1%; Score 86; DB 16; Length 303;
Best Local Similarity 22.6%; Pred. No. 37;
Matches 28; Conservative 19; Mismatches 39; Indels 38; Gaps 4;

QY 24 KITEINKAIDDAIAIEQSETIDPMKVPDHAKFERHGVDFKGLAMRNIEARGLK 83
DB 71 KVLDEARLADDA-----BHG-----ELRGSRLVTTQEGELRQ 106
QY 84 MKRGGANVKGEGIVYKAHLIGV-----HDDIVSMEDLAKGLDHPHTTVISDIQ 136
DB 107 L-----VPALQAFARLHALQVQLSTSLHLDLIGERPDVALIRGLRLEDSTYHNAYQLA 159
QY 137 DFVY 140
DB 160 SFEV 163

RESULT 24
P71763 PRELIMINARY; PRT; 339 AA.
AC P71763;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical protein Rv1482c.
GN Rv1482C OR MT1529 OR MTC1277.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwin M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: SOME, TO THE GRPE FAMILY.
DR EMBL: Z79701; CAB02032.1; -.
DR EMBL: AE007022; AAK45794.1; ALT_INIT.
DR TIGR: MT1529; -.
DR TubercuList; Rv1482c; -.
KW Hypothetical protein; Complete proteome.

```

```

SO SEQUENCE 339 AA; 38290 MW; A0526C082B219D18 CRC64;
Query Match 8.1%; Score 86; DB 16; Length 339;
Best Local Similarity 28.3%; Pred. NO. 42;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;
QY 33 IDDATAIEQSTIDTPMNVPHADKFFERVIVDPKFGELAMNIEARGIKOKRR----- 86
DB 187 LDPAAVAADALIQATDCLKADYDEPLEIERGRGRKMAARAALDLVDGASQSPRETWLRLL 246
QY 87 -----QGDANYKGEQGIYKAHLIGVHDDIVSMEXDLYAKKLDLPPT--HVISDI 135
DB 247 LIRAGFPFRPQGIANVRNEMGVAEHALDLMGKODIKYAAEYD-----GDHHLTSRYHTRKDI 301
RESULT 25
006166
AC 006166: PRELIMINARY; PRT: 1661 AA.
ID 006166:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Mature PARASITE-infected erythrocyte surface antigen (Antigenic
protein PfEMP2).
OS Plasmodium falciparum.
OC Eukaryota, Alveolata, Apicomplexa, Haemosporida, Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO;
RX MEDLINE=92158014; Pubmed-1741020;
RA Coppel R.L.;
RT "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT binds human erythrocyte protein 4.1."
RL Mol. Biochem. Parasitol. 50:335-347(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO;
RX Kun J.F., Waller K.L., Coppel R.L.;
RA Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
[3]
RP SEQUENCE OF 797-850 FROM N.A.
RC MEDLINE=9112844; Pubmed-1478701;
RX Saul A., Yeganeh F., Howard R.J.;
RA "Conservation of repeating structures in the PfEMP2/MESA protein of
RT Plasmodium falciparum."
RL Immunol. Cell Biol. 70:353-355(1992).
DR EMBL: S52458; AAB24869.1; -
DR InterPro: IPR001623; DnaJ.N.
DR Pfam: PF00226; DnaJ. 1.
DR SMART: SM00271; DnaJ. 1.
DR PROSITE: PSS0076; DnaJ_2. 1.
KW Antigen.
SO SEQUENCE 1661 AA; 195479 MW; AF340527D85A9D29 CRC64;
Query Match 8.1%; Score 86; DB 5; Length 1661;
Best Local Similarity 20.7%; Pred. NO. 3.5e+02;
Matches 46; Conservative 41; Mismatches 67; Indels 68; Gaps 10;
QY 15 VSADPIHDKITTEINKAIDAIAIEQSETIDPMKVP-----DHADKFERHVGIVDEKG 69
DB 604 VKGEPIIEEVEKEIKKQVEDGIKE-NDTEGNDKVKGEPIIEEVEKEIKKHV----- 655
QY 70 ELAMNIEARGIKOKRGDANVKEEGI-----VKAHLIGVHD-----D 110
DB 656 -----EEGKKNNTGNDKVKGEPIITEEVEKEIKKHVEGKIKENDTESKDVIGOE 707
QY 111 IVSMEDLAYKIGDHPPTHVIVSIDQDPEVVALSLIEISDEGNITMTSFVROPFANYNHIG 170
DB 708 IITEEVEKGIKEND-----TEKKDKVIGDEM-ITEE-----VKKEIKQEGKGNKEN 753
QY 171 GLSIIDPIFGVLSVDVLAIFQDVTVRKEMTKVLAPAFKRELEK 212

```

Db	754	ILEIKDIYIG-----QEVITEEVKKVI-----KKKVEK	781
Db	754	ILEIKDIYIG-----QEVITEEVKKVI-----KKKVEK	781
RESULT	26		
O91Z50			
AD	O91Z50	PRELIMINARY;	PRF: 1154 AA.
AC	O91Z50;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DE	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	MASS1.3.		
GN	MASS1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6;		
RA	MEDLINE=21432018; PubMed=11545713;		
RA	Skradski S.L., Clark A.M., Jiang H., White H.S., Fu Y., Ptacek L.J.;		
RT	"A novel gene causing a mendelian audiogenic mouse epilepsy."		
RL	Neuron 31:537-544(2001).		
DR	EMBL; AF405694; AAL06014.1; -		
DR	MGI; MGI:1274784; Mass1.		
DR	InterPro; IPR003644; Calx_beta.		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	Pfam; PF03160; Calx-beta; 6.		
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.		
SO	SEQUENCE 1154 AA; 123814 MW; CBDACF4ECE3B5D06 CRC64;		
Query Match	8.0%; Score 85.5; DB 11; Length 1154;		
Best Local Similarity	21.4%; Pred. No. 2.4e+02;		
Matches	41; Conservative 32; Mismatches 78; Indels 41; Gaps		
OY	26	TEEINKAIDDAIAIEQSEFIDPMKVPDPHADKFERHV-----GIYDFKGLAMRNIEA	78
Db	882	TVTVAIILNDVNVAGIVSFOTASRSVIGHEGEMIDRFHVRTFPGRGNTVNMKVVGQNDLV	941
OY	79	RGIKMKRKGQGANVKE---EGIVKAHLILGVHDDIVSM-----YDLAVKLDGLH	126
Db	942	-----NFRNFTGQLFFSEGLINKTIFVHLLDNDIPEEKEVYQVVLVDV--KTQGS	990
OY	127	PTTHVSDIDQFVVALSLSEISDEG-----NITMTS-FEYRQFANV-----NHIGLSIL	175
Db	991	PAGVALLDAQGYAANLVTEASDEPRGCVLNFALSSFFVLQEAENVITQLFVNRPFOSLCAI	1050
OY	176	DPFEGVLSVLT 187	
Db	1051	NVTATVPGIVS 1062	
RESULT	27		
O60310			
ID	O60310	PRELIMINARY;	PRF: 1441 AA.
AC	O60310;		
DT	01-AUG-1998 (TREMBLrel. 07, Created)		
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	K1AA0564 protein (Fragment).		
GN	K1AA0564.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRIN;		
RA	MEDLINE=98290545; PubMed=9628581;		
RA	Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,		
RT	Ohara O.;		
TX	"Partial O. of the coding sequences of unidentified human genes. IX.		

RA McMillan D.R., Kates-Wandover K.M., Richardson J.A., White P.C.;
 RT "Very Large G protein-coupled Receptor-1, the Largest Known Cell
 RT Surface Protein, is Highly Expressed in the Developing Central Nervous
 RT System";
 RL J. Biol. Chem. 277:785-792(2002).
 DR EMBL: AF435926; AAL30812.1; -;
 DR MGD: MGI:1274784; Mass1.
 DR InterPro: IPR003644; Calx_beta.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR000832; GPCR_secretin.
 DR InterPro: IPR000203; PKD_cys_rich.
 DR Pfam: PF00002; 7tm_2; 1.
 DR Pfam: PF03160; Calx_beta; 19.
 DR SMART: SM00237; Calx_beta; 20.
 DR PROSITE: PS50221; GPS; 1.
 DR PROSITE: PS50261; G_PROTEIN_RECEP_F2_4; 1.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; UNKNOWN; 1.
 DR Receptor.
 SQ SEQUENCE 6298 AA; 687437 MW; F99E731B0ADB66D CRC64;

Query Match 8.0%; Score 85.5; DB 11; Length 6298;
 Best Local Similarity 21.4%; Pred. No. 2.2e+03;
 Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

OY 26 TEEINKAIDDAIAIEQSEITIDPMKVPDHADEKFEHRY-----GIYDFKELAMRNEA 78
 DB 2686 TVTVNLANDNVAGIVSFGASRSVIGHEGEMLOFHVHRTPPRGAVTVMKKVGVGNLEV 2745
 OY 79 RGLKQKROGDANVGE---EGIVKAHLIGVHDDIVSKE-----YDLAYKGLDLH 126
 DB 2746 -----NFAFNGTQLFSECTLNKTFVHLLDNIPEKEVGVVLDV--KTQGV 2794
 OY 127 PTHVTSIDQDFVVALSLEISDEG-----NITMTS-FEVRQFANV-----NHIGSL 175
 DB 2795 PACVALLDKGGTAAVLFEASDEPHGVNLALSSRFVLDQAVNTTQLFVNRREGSLGAI 2854
 OY 176 DPFGVLSVLT 187
 DB 2855 NVTYATVPGLVS 2866

RESULT 31
 OY 2686 TEEINKAIDDAIAIEQSEITIDPMKVPDHADEKFEHRY-----GIYDFKELAMRNEA 78
 ID 09UZS6 PRELIMINARY; PRT; 292 AA.
 AC 09UZS6:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein PAB0711.
 GN PAB0711.
 OS Pyrococcus abyssi
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORSAY;
 RA Hellig R.;
 RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
 RT structure and evolution.";
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ248286; CAB49980.1; -;
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 292 AA; 31578 MW; AE410B3578BBFA8C CRC64;

Query Match 8.0%; Score 85; DB 17; Length 292;
 Best Local Similarity 20.7%; Pred. No. 42;
 Matches 44; Conservative 31; Mismatches 62; Indels 76; Gaps 7;

OY 29 INKAID-----AIAIEQSEITIDPMKVPDHADEKFEHRYGIVDF----- 68
 DB 11 INKVIENKVIIPPLIVLIFSVLAFLERESNIKPLNEEAGVITIEKHAISDMKLPNL 70

OY 69 -----GELAMRNEA-----RGLKOMKROGDANVKGEEGIVKAHLIGVHDDIVSMEY 116
 DB 71 KLVLGLGLQLLLSNAVQSIHHVKTGSTMGAFKLGLENVIOMFL----- 118
 OY 117 DLAYKGLDLHPTTHVTSIDQDFVVALSLEISDEGNTMTSFEVRQF-ANVNVNIGGL--S 173
 DB 119 -----NVISTL-----TVLVAFLIAPPMIIVIGGILES 149
 OY 174 ILDPFGVLSVLTALFQDTPVRKEMTKVLAPAF 206
 DB 150 TGAIVTIGILVLTSLGLIGSFAIGMTSVIVPAY 182

RESULT 32
 OY 174 ILDPFGVLSVLTALFQDTPVRKEMTKVLAPAF 206
 ID 08RDZ2 PRELIMINARY; PRT; 401 AA.
 AC 08RDZ2:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ABC transporter permease protein.
 GN FNL349.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhatlacharya A., Bartman A., Gardner W., Greshkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Golsman E., Bernal A.,
 RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
 RA Fongstein M., Kyrpides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL: AE010639; AAL95545.1; -;
 KW Complete proteome.
 SQ SEQUENCE 401 AA; 44082 MW; AE17A5DAEBF49B2 CRC64;

Query Match 8.0%; Score 85; DB 16; Length 401;
 Best Local Similarity 23.2%; Pred. No. 64;
 Matches 52; Conservative 40; Mismatches 72; Indels 60; Gaps 13;

OY 13 VAVSADPIHYDKITTEINKAIDDAIAI-----ESETI-----DPMKVPDHADEKFEHRY 62
 DB 126 IDIDTFLIYPMYTHNIDKELKDEAIVGSHVYGEKGFVHFNEELIKIVGRL---KQT 181
 OY 63 GIYDFKELAMRNEAIEARGLKOMKROGDANVKGEEGIVKAHLIGVHDDIVSMEYDLAYKL 122
 DB 182 GI-GFATVAVFNONTAKOLAKASERITANKVAEEDVSS-VMIKAPGVDSK--LASK- 236
 OY 123 GDLHPTTHVTSIDQDFVVALSLEISDEGNTMTSFEVRQFANVNH-----IG 170
 DB 237 -----ISKELSKEGIFAMTS---KKRVNMISSNKLKVLSTSLVIG 274
 OY 171 GLSIDPFGVLSVLTALFQDTPVRKEMTKVLAPAF--REL 210
 DB 275 AIVLSIV--VLSISFTAVFNER-KKEMAVLRVLGASKMLREI 315

RESULT 33
 OY 171 GLSIDPFGVLSVLTALFQDTPVRKEMTKVLAPAF--REL 210
 ID 09CFU9 PRELIMINARY; PRT; 878 AA.
 AC 09CFU9:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Cation-transporting ATPase.
 GN Y0AB OR L13366.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=111403.
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis 111403."
RL Genome Res. 11:731-753(2001).
DR EMBL: AEO06368; AAK05464.1; -.
DR HSSP: P04191; 1E0U.
DR InterPro: IPR001757; ATPase_E1-E2.
DR InterPro: IPR004014; Cation_ATPase.
DR InterPro: IPR000661; H/K_Na/K_ATPase.
DR InterPro: IPR001454; Hlgase/hydlase.
DR Pfam: PF00689; Cation_ATPase_C; 1.
DR Pfam: PF00690; Cation_ATPase_N; 1.
DR Pfam: PF00122; E1-E2_ATPase; 1.
DR Pfam: PF00702; Hydrolase; 1.
DR PRINTS: PR00119; CATAPASE.
DR PROSITE: PS00154; ATPASE_E1_E2; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 878 AA; 96032 MW; 0D322B263E3DF36 CRC64;

Query Match 8.0%; Score 85; DB 16; Length 878;
Best Local Similarity 24.7%; Pred. No. 1.8e+02;
Matches 39; Conservative 33; Mismatches 44; Indels 42; Gaps 8;

QY 72 AMRNIEARGLKQKRGDANKGEGIVKA-HLLIGVHDDISMEYDLAKGLDHPPTH 130
DB 106 ALRMSAPSAKVLK-----NGEKTSIPARELYVG--DIVSL-----AGDFIPADG 149
QY 131 VISIOGFVALSLSEISDEGNITMTSEVROFANV-----NHIGSLIDLPFG 180
DB 150 RLIDVQMLRV-----EEGMLTGESEPEVEKESVDVIGSEVALGDRKNVSSSLV--VYG 200
QY 181 VLSDVLTATFODTVRKEMTKVLAPA-----FKRELEK 212
DB 201 RADFLVTAIAEQTEIGKIAQLLEFAKQKPLQOKLEK 238

RESULT 34
Q8RQU4 PRELIMINARY; PRT; 334 AA.
ID 08RQU4
AC 08RQU4:
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Homoisocitrate dehydrogenase.
GN HICD.
OS Thermus thermophilus.
OC Bacteria; Thermus/deinococcus group; Deinococci; Thermales;
OC Thermaceae; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyazaki J., Kobashi N., Nishiyama M., Yamane H.;
RT "Characterization of homoisocitrate dehydrogenase from Thermus
RT thermophilus HB27."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB075751; BAB88861.1; -.
SQ SEQUENCE 334 AA; 35922 MW; 82B018FED744FB49 CRC64;

Query Match 7.9%; Score 84.5; DB 2; Length 334;
Best Local Similarity 22.8%; Pred. NO. 55;
Matches 50; Conservative 34; Mismatches 94; Indels 41; Gaps 7;

QY 10 VAFVAVSADPIHYDKITEINKAIDDAIALEQSETIDPMKVPDPAHAKFEHVGIVDFKG 69
DB 114 LVIVRENTGILYEOERRRDLVAIDAVISKASERI-----G 151

QY 70 ELAMRNIEARGLKQKRGDANKV-----KGEISYKAHLLIGVHDDIV--SMEYD 117
DB 152 RAALRIAEGRPRKTLIHAKNAVPLTQGLFDLVKFAKADFLVNVQDITIVNCAKQV 211
QY 118 LAYKLGDLHPTTHVISDI-ODFVAL--SLEISDEGNITMTSEVROFANVNHIGSLI 174
DB 212 MRPERFVIVTNTNLGILSLDLAAGLVGLGLAVSGNIGDTTAVFEPVHGSADPIAGKI 271
QY 175 LDPFVGLSDVLTATFODTVRKEMTKVLAPAFKRELEK 212
DB 272 ANPTAALIS---AAMMDIYGEKKAARKVAKVDVLER 307

RESULT 35
Q9ALA9 PRELIMINARY; PRT; 547 AA.
ID Q9ALA9
AC Q9ALA9:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RA Wong H.-C., Lu K.-H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AY017169; AAG48876.1; -.
DR HSSP: P06139; 1GRL.
DR InterPro: IPR001844; Chaperon_Cpn60.
DR InterPro: IPR002423; Cpn60_TCP1.
DR Pfam: PF00118; Cpn60_TCP1; 1.
DR PRINTS: PR00296; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KM ATP-binding; Chaperone.
SQ SEQUENCE 547 AA; 57666 MW; 4A999BAF55BEC143 CRC64;

Query Match 7.9%; Score 84.5; DB 2; Length 547;
Best Local Similarity 19.7%; Pred. NO. 1.1e+02;
Matches 47; Conservative 43; Mismatches 76; Indels 73; Gaps 11;

QY 13 VAVSADPIHYDKITEINKAIDDAIALEQSETIDPMKV--PDHAKFE-RHVGIVDFKG 69
DB 107 VVAGMNPV-----DLKRGIDKAVAA---AVEELKAMSKDCSTSTELQVGTISANS 154
QY 70 ELAMRNIEARGLKQKRGDANKGEGIVKAHLLIGVHDD--IVSMEDYLA----- 120
DB 155 DSSVGKIIAEAMEKVGDRGVTV--EEG-----QALHDLVDVEGMDGRLGSPYFI 205
QY 121 -----KLGDLHPTTHVISDIODFVALSLSEISDEGNITMTSF 157
DB 206 NNQSSGVSELPFLVLDKKSINRELPLALEVAAKSRLLIADVEDALATL-- 262
QY 158 EVROFANVNHIGL---SLIDPFGVLDVLTATFODTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VVNNMRGIVIVAIAVAFGFG--DRRAMLQDIALTLGNGVISEVGLLELEK 311

RESULT 36
Q8SXM3 PRELIMINARY; PRT; 1429 AA.
ID 08SXM3
AC 08SXM3:
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE LD43558p.
GN RPII215.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarini H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celiker S.;
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
RL EMBL: AY089555; AAL90293.1;
SQ SEQUENCE 1429 AA; 157524 MW; 540A045520CAEDA6 CRC64;

Query Match 7.9%; Score 84.5; DB 5; Length 1429;
Best local Similarity 22.7%; Pred. No. 3.8e+02;
Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12;

OY 17 ADPIHYDKITEEINKAIDAIAIEOSEFTIDPMKVPDHA--DKFERHVGIV-----DFKG 69
DB 228 ADPOTNEIQOAIKKAKDVINYIOKAHNELEPTGNTLRQTFENKVRNILDARDKTG 287
OY 70 ELAMRNI-EARGLKQMKRO-----GDANYGEE---GIYKAHLIGVHD 109
DB 288 GSAKSKLTFEYNNLKAMVSGSKGSNINISQVACVGOQNEGRIPYGFRRKRLPHFIKD 347
OY 110 D-----IYSKMYDLAYKLGDLHPT---HVISDIQDFVALSLSESGNI-----T 153
DB 348 DGPESRGFEVNSY-----LAGLPSEFYFHANGG-REGLDIYAVTAETGYIQRLIKA 401
OY 154 MTSFEYROPANVNHIGLSIL---DPIFGVLSL----- 184
DB 402 MESVWVNYGIVANSYGLQILRYGEGDGLGELVEFQNPVYLSKNSPKRKRPFMSNE 461
OY 185 -VLTALFDPTVRKEMT 199
DB 462 RLAKKVFETDVIKEMT 477

RESULT 37

O99XZ07 PRELIMINARY; PRT; 314 AA.

AC O99XZ07;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein TM0800.
GN TM0800.

OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=23356;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS8 / DSM 3109;
RC MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima";
RL Nature 399:323-329 (1999).
DR EMBL: AE001747; AAD35882.1; -.
DR HSSP: P12268; 1B30.
DR TIGR: TM0800; -.

DR InterPro: IPR004136; 2nprop_dioxygen.
DR InterPro: IPR003009; FMN_enzyme.
DR Pfam: PF03060; NPD_1.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 314 AA; 33673 MW; 3321EBCF7647D9B8 CRC64;

Query Match 7.9%; Score 84; DB 16; Length 314;
Best local Similarity 23.5%; Pred. No. 56;
Matches 48; Conservative 32; Mismatches 76; Indels 48; Gaps 10;

OY 23 DKITEEINKAIDAIAA--IEOSEFTIDPMKVPDHA--DKFERHVGIV-----GIVDFKGEIA 72
DB 122 DSLARVERAGADVAIEGMEGSGHIGEVTFVLVKNKSRVNIPIYAGIADGRGMAA 181
OY 73 MRNIEARGLKQMKR-----GDANYGEEGVYKAHLIGVHDIVSEMDLAKKLDLP 127
DB 182 AFALGAEAAYOMGTRFVASVESDVHPYKKEIVKA---SIRDIYVY-----GAKLD--HP 230
OY 128 T-----THVISDIQDFVALSLSEISD-----EGNTMTSFEYROPANVNHIGL 172
DB 231 ARVLRTPPARKIDEMEFENPMQAEMLVSLRAVYEGDLERGSFVYGOSAGLIDEL--- 287
OY 173 SLIDPIFGVLSVLTFAIFDPTVRK 196
DB 288 ---KPVKQIILEDLKE-FKRETVK 307

RESULT 38

O99X76 PRELIMINARY; PRT; 392 AA.

AC O99X76;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Phosphoenolpyruvate carboxylase (EC 5.4.2.7).
GN DRM OR SAV0139 OR SA0134.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilliales;
OC Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);
RC MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus";
RL Lancet 357:1225-1240 (2001).
DR EMBL: AP003358; BAB56301.1; -.
DR EMBL: AP003129; BAB41354.1; -.
DR InterPro: IPR002599; Metalloenzyme.
DR Pfam: PF01676; Metalloenzyme: 1.
RT Isomerase; Complete proteome.
SQ SEQUENCE 392 AA; 43795 MW; 6012D71C9DBD549 CRC64;

Query Match 7.9%; Score 84; DB 16; Length 392;
Best local Similarity 24.1%; Pred. No. 75;
Matches 45; Conservative 35; Mismatches 85; Indels 22; Gaps 10;

OY 14 AVSADPIHYDKITEEINKAIDAIAIEOSEFTIDPMKVPDHA--DKFERHVGIVDFKGEIA 72
DB 65 AVDQPEAYITKLSAASVGDKTMTGHNEMGLNIMQFPKVPYVNPFPRELIQOIEKMGKRV 124
OY 73 MRNIEARGLKQMKRQGDANYGEEGVYKAH---LLIGVHDIVSME--YDLAYKGLDL- 125
DB 125 VANKPASGTQIIDEMGEHQKGTGLIVYTSADPVLGIAAHEDIIIPLEIYDICEKVELH 184

RESULT 41

09CH23 PRELIMINARY: PRT: 775 AA.

ID 09CH23 AC 09CH23

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)

DE Cation-transporting ATPase.

GN YFGG OR LD0575.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Lactococcus.

OX NCBI_TaxID=1360;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IL1403;

RX MEDLINE=21235186; PubMed=11337471;

RA Bolotin A., Winkler P., Manger S., Jallion O., Malarme K.,

RA Weissenbach J., Ehrlich S.D., Sorokin A.;

RT "The complete genome sequence of the lactic acid bacterium Lactococcus

RT lactis ssp. lactis IL1403."

RL Genome Res. 11:731-753(2001).

DR EMBL: AE006291; AAK04673.1; -

DR InterPro: IPR001757; ATPase_E1-E2.

DR InterPro: IPR001454; Hlgase/hydase.

DR Pfam: PF00122; E1-E2_ATPase.1.

DR Pfam: PF00702; Hydrolase.1.

DR PRINTS: PR00119; CATATPASE.

DR PROSITE: PS00154; ATPASE_E1_E2; UNKNOWN_1.

KW Complete proteome.

SQ SEQUENCE 775 AA: 85467 MW: 84D5E84E49A833BE CRC64;

Query Match 7.9%; Score 84; DB 16; Length 775;

Best Local Similarity 19.8%; Pred. No. 1.8e+02;

Matches 50; Conservative 37; Mismatches 103; Indels 62; Gaps 6;

QY 22 YDKIREIKALIDDAIALEOSETIDPMKVPFHADKFERHVGIVDKGEL-----AMRN 75

DB 397 YKGVSEIHLLEEGFRVYLAKTKE--KIYDQNLGAVALGVLANPFIENAKSTFNV 454

QY 76 IEARKLQKMRQD-----ANVGESEIVKAHL-----104

DB 455 FAEQGVNIVISGDNQGTSAVAKRAGITGAERFDANLTKREDLQAVESYTVGRVT 514

QY 105 -----IGVHDDIVSEYD-----LAYKLGLDHPHTHVISDIQDFVALSLSEISDE 149

DB 515 PQKRLVQALKRKDHTVAMTGDGVNDILAMKSADCSIMASGSDAATVAGVLLDSDF 574

QY 150 GNITMSEFVROFANVNVNIGGSLIDPIFGVLSVDLTAIPQDTVRKEMTK-----V 201

DB 575 GIMTOVVTGKRVNNAVQSAILELVKNLFSIIILAIISAFVFTYPLQASQSLISLFTI 634

QY 202 LAPAFKRELEKN 213

DB 635 GIPGFLLSLEEN 646

RESULT 42

093NX9 PRELIMINARY: PRT: 9510 AA.

ID 093NX9 AC 093NX9

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)

DE AmpH1.

GN AmpH1.

OS Streptomyces nodosus.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomyces; Streptomyces; Streptomyces.

OX NCBI_TaxID=40318;

RN [1]

RP SEQUENCE FROM N.A.

RA Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., O'Leary M.;

RT "The amphothericin biosynthetic gene cluster from Streptomyces

RT nodosus."

RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF357202; AAK73501.1; -

DR InterPro: IPR001227; AC_transferase.

DR InterPro: IPR002198; ADH_short.

DR InterPro: IPR001899; FAD.

DR InterPro: IPR00794; Gram_pos_anchor.

DR InterPro: IPR001005; MyD_DNA_binding.

DR InterPro: IPR003880; Pantane_attach.

DR Pfam: PF00698; Acyl_transf. 6.

DR Pfam: PF00106; adh_short.1.

DR Pfam: PF00109; ketoacyl-synt. 6.

DR Pfam: PF02801; ketoacyl-synt_C. 6.

DR Pfam: PF00550; pp-binding. 6.

DR TIGRfams: TIGR00128; fadD. 6.

DR PROSITE: PS50075; ACP_DOMAIN. 6.

DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.

DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE: PS00037; MYD_1; UNKNOWN_1.

DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_6.

KW Phosphopantetheine.

SQ SEQUENCE 9510 AA: 992817 MW: 1A8E003A15D478CF CRC64;

Query Match 7.9%; Score 84; DB 2; Length 9510;

Best Local Similarity 24.5%; Pred. No. 5.1e+03;

Matches 53; Conservative 39; Mismatches 80; Indels 44; Gaps 10;

QY 3 KFLIAVAVFVAVSADPHYD---KIT--EELNK-----AIDDAIALEOGE 44

DB 582 REPVEFAALDADLADATPILHDPLKRVLNGEDADRLDRETYAQPALFAVEVALYRLLESEF 641

QY 45 TIDPMKVPDHA--DKFERHVGIVDFKGLAMRNIEARG-LKQKROGDANVGESEIVKA 101

DB 642 EVKPDHLGHSVGEIAAAHVAGV-FSLDDAATLVAAARGRLQALPFGAMVAVQASEDEV 700

QY 102 HLLIGVHDDIVSEYDLAKKLDLHPTTHVISDIQDFVALSLSEISDEGNIT-----153

DB 701 APLLAGHEDLV---LAAVNG---PSAVVLSGDETVTVELAARLAADGRKTSRLRVSHA 753

QY 154 ---MTSEFVROFANVNVNIGGSLIDPIFGVLSVDV 185

DB 754 FHSPLMAPMLDEFRRVVE---GLTLHSPILPVVSDV 786

RESULT 43

093R44 PRELIMINARY: PRT: 311 AA.

ID 093R44 AC 093R44

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)

DE HSP60 (Fragment).

GN HSP60.

OS Yersinia pseudotuberculosis.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Yersinia.

OX NCBI_TaxID=633;

RN [1]

RP SEQUENCE FROM N.A.

RA Yang R.F.;

RT Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Du Z.M.;

RT Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AB064590; BAB61891.1; -

DR InterPro: IPR002423; Cpn60/TCP-1.

DR Pfam: PF00118; cpn60_TCP1.1.

Query Match	7.88;	Score 83.5;	DB 2;	Length 329;
Best Local Similarity	18.18;	Pred. No. 65;		
Matches 43;	Conservative 44;	Mismatches 81;	Indels 69;	Gaps 9

RESULT 47	
09C0B4	
ID 09C0B4	PRELIMINARY; PRT; 360 AA

SEQUENCE	360 AA;	42730 MW;	87934E92EB5EF926 CRC64;
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

QY	21	HYDKITEELINKALDDAIAAEOSERTIDPPKAVPRPHADKFERHNGIVDFKELAMRNTEAR	79
Query Match	7.8%	Score 83.5	DB 4; Length 360;
Best Local Similarity	21.7%	Pred. No. 74;	
Matches	31; Conservative	29; Mismatches	66; Indels 17; Gaps 4

RESULT 48
Q97K37

Query Match	7.88	Score 83.5	DB 16	Length 474
Best Local Similarity	23.38	Pred. NO. 1.1e+02		
Matches 34	Conservative 22	Mismatches 65	Indels 25	Gaps 3

Db 175 YRCVHPPGICKIKKYLANHIVNLA 200

RESULT 49		
Q82IY3		
ID	Q82IY3	PRELIMINARY; PRT; 548 AA.
AC	Q82IY3;	
DT	01-MAR-2002 (TREMBlrel. 20, Created)	
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)	
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)	
DE	60 kDa chaperonin.	
GN	GROEL OR YPO0351.	
OS	Yersinia pestis.	
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	

Query Match	7.8%;	Score 83.5;	DB 16;	Length 548;
Best Local Similarity	18.1%;	Pred. No. 1.3e+02;		
Matches 43;	Conservative 45;	Mismatches 80;	Indels 69;	Gaps 9

RESULT 50	
Q9X6Y3	
ID Q9X6Y3	PRELIMINARY;
NO Q9X6Y3	PRT; 544 AA

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE 60 kDa chaperonin (Protein cpn60) (GroEL protein).
OS GROEL.
OS Bacteroides forsythus.
OC Bacteria; CFb group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=28112;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43037;
RX MEDLINE=99452396; PubMed=10524765;
RA Reid H.T., Rigallo M.P.,
RT "Identification and nucleotide sequence of the heat shock protein 60
RT (GroEL) gene of Bacteroides forsythus.";
RL DNA Seq. 9:359-364(1998).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

Query Match	7.8%;	Score 83;	DB 2;	Length 544;
Best Local Similarity	22.3%;	Pred. No. 1.4e+02;		
Matches 47;	Conservative 36;	Mismatches 72;	Indels 56;	Gaps 11

Search completed: February 21, 2003, 10:09:18
Job time : 104 secs